THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

Vol. 65, Part 2 1 February 1952 No. 386 B

A Coil System for an Air-cored Betatron

By R. LATHAM, M. J. PENTZ and M. BLACKMAN

Department of Physics, Imperial College, London

MS. received 13*th July* 1951, *and in amended form* 19*th November* 1951

ABSTRACT. The paper describes the design of an air-cored coil system for a small betatron and measurements of the field configuration produced by this system.

§1. INTRODUCTION

This work arose from an attempt to construct a small betatron to accelerate electrons to an energy of 300 kev in the presence of a gas pressure of the order of 10^{-3} mm Hg. This is a factor of about one thousand larger than is usual in a conventional betatron and makes it necessary to have a very short acceleration time of the order of 5 microseconds. It is only possible to do this with an air-cored coil, the current through which was provided in our case by a condenser discharge.

Many difficulties were encountered in the course of the experiments, and it has now been decided to discontinue them. A coil system which produced a magnetic field of the right shape for betatron acceleration was, however, devised and tested. It is the purpose of this paper to give details of the coil system, with a summary and discussion of the tests carried out.

§2. DETAILS OF THE PROBLEM

The magnetic field required for the particular purpose of these experiments had to satisfy two main conditions. Firstly, as mentioned above, it had to rise very rapidly with time, which precluded the use of an iron core. Secondly, a steady magnetic field of the same shape as the betatron accelerating field should remain after the end of the acceleration, i.e. after the condenser discharge had ceased. In fact a steady field was not used, but the peak of a 50 c/s a.c. field was sufficiently uniform to be considered as a steady field for the millisecond or so during which electrons were expected to circulate after the accelerations.

The cycle of events was therefore as shown in fig. 1. At the peak of the 50 c/s current in the coils, a critically damped condenser discharge was arranged which temporarily reduced the current to zero or to a slightly negative value. This was followed by the decay of the condenser current, and consequent rise of the total current to its original steady value. An electron injected at A would be accelerated over the region AB and continue to circulate over the region BC.

§ 3. COIL SYSTEM: THEORETICAL DESIGN

The simplest kind of coil system to consider would be a pair of coaxial coils, but this gives a field varying much more rapidly with radius than the $1/r$ limit for radial stability. Two coaxial pairs of suitable dimensions carrying different currents can give the correct dependence on r but only over a limited region of r. With three pairs of coils the system is more flexible and a suitable field can be achieved over a reasonable volume between the coils. The calculations are simplified by using the tables of H_z and H_ρ given by Blewett (1947) for circular coils.

A satisfactory coil system was first found by trial and error, and optimized by varying the parameters systematically. This had the dimensions and currents shown in fig. 2 and gave a field shape in the central plane $z=0$ as in the broken curve in fig. 6.

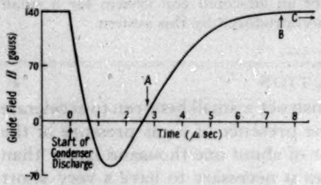


Fig. 1. Time variation of guide field.

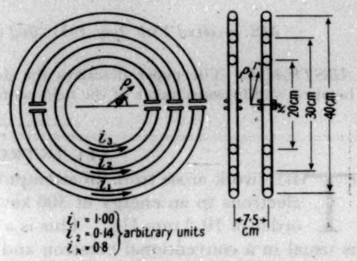


Fig. 2. Dimensions of coil system.
(In the above figure the innermost coil should be labelled i_1 and the outermost i_3.)

A detailed study was next made of the field components H_z and H_ρ throughout the region between the coils. The field components at a point are given by the following relations (Blewett 1947).

$$\frac{H_z}{I}=\frac{2}{[(a+\rho)^2+z^2]^{1/2}}\left[K(k)+\frac{a^2-\rho^2-z^2}{(a-\rho)^2+z^2}E(k)\right],$$

$$\frac{H_\rho}{I}=\frac{2z}{\rho[(a+\rho)^2+z^2]^{1/2}}\left[-K(k)+\frac{a^2+\rho^2+z^2}{(a-\rho)^2+z^2}E(k)\right],$$

where E and K are the standard elliptic integrals, a is the radius of a given coil, ρ and z the coordinates of the point with respect to the centre of the coil and $k=4a\rho/[(a+\rho)^2+z^2]^{1/2}$.

The fields are calculated for each of the six coils and added. Calculations were carried out over four planes at $z=0$, 0·8, 1·2, and 1·8 cm and the results are shown graphically in figs. 3 and 4.

From the above it was then possible to construct the lines of force in the region between the coils to be occupied by the doughnut and to deduce the region of axial stability. The usual stability condition requiring the axial magnetic field to decrease with radius refers to motion near the median plane. For other planes a more general criterion is required, namely that, with H_z positive, H_ρ and z should have opposite signs. The region of radial stability can be found from the H_z curves by choosing the region over which H_z varies less rapidly than $1/r$. This is shown in fig. 5, the regions of radial and axial instability being shaded, and labelled r and z respectively. The doughnut cross section is shown by the circle marked D and an effective doughnut cross section by the circle marked d. It will be seen that a reasonably large doughnut cross section can be accommodated within the stable region.

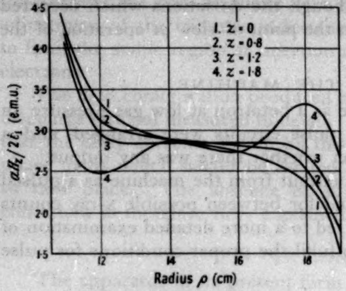


Fig. 3. Calculated variation of H_z with ρ, off median plane, where z (cm) is the distance from the median plane.

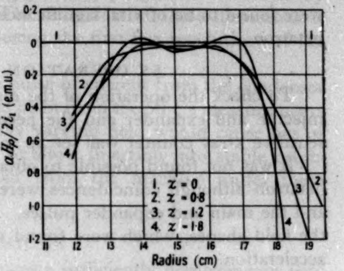


Fig. 4. Calculated variation of H_ϱ with ρ, off median plane, where z (cm) is the distance from the median plane.

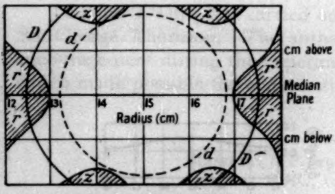


Fig. 5. Cross section of region of stability. Shaded regions are regions of radial, r, and axial, z, instability. D indicates the actual and d the effective circular doughnut cross section.

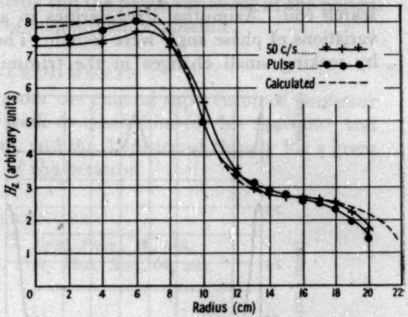


Fig. 6. Measured and calculated variation of H_z with ρ.

§4. COIL SYSTEM: EXPERIMENTAL DESIGN

The magnitude of the currents required (1·0, 0·14 and 0·80 in arbitrary units) suggested a series–parallel arrangement, with the middle and outer coils in parallel, connected in series with the inner coil. Two additional pairs of trimming inductances were necessary to obtain the required current ratios in the coils. The requirement that the distribution of currents should not be different for the condenser discharge current and the 50 c/s current necessitated that the resistance ratios be the same as the inductance ratios and decided the diameters of the material used to form the coils.

For electron energies of 300 kev the steady magnetic field at the orbit had to be of the order of 150 gauss after the acceleration, requiring a current of 5 000 amp (2 500 amp in each set of coils). Such currents were easily obtainable from a welding transformer.

To prevent the condenser discharge affecting the welding transformer a simple filter circuit was incorporated in the connections between the transformer and the main coils. Initial tests were made on this coil system under full load conditions by measuring the 50 c/s field and the pulse field separately. In both cases the calculated variation of field on the $z=0$ plane with ρ was found to be close to that actually measured (fig. 6), though the deviations which occurred were found to be of vital significance from the point of view of operation of the betatron.

§5. OPERATION OF THE MACHINE

To check the operation of the machine as a betatron at low gas pressure an injector and expander and the necessary pulse circuits were arranged, and a sensitive x-ray counter was set up to detect whether there was any output.

It was not found possible to obtain an output from the machine as a pulsed betatron although coincidences were looked for between possible x-ray counts and the main and expander pulses. This led to a more detailed examination of the field shapes, which were found not to fulfil the proper conditions for pulse acceleration.

§6. DETAILED FIELD MEASUREMENTS

Measurements were made at 50 c/s by a null method using a small exploring search coil. Azimuthal irregularities in amplitude and phase angle and radial variations of phase angle were found. These defects were considerably reduced by making small changes in the trimming inductances and by altering the physical arrangement of the connecting leads to the coils to minimize stray fields. The final field shape obtained was as shown in fig. 7. This represents the best which could be done with the coil system as at present designed, the irregularities being at the points where the current enters or leaves the main coils. The effect of these asymmetries was calculated using the work of Goward (1948) and assuming the apparatus to be run as a straightforward 50 c/s betatron.

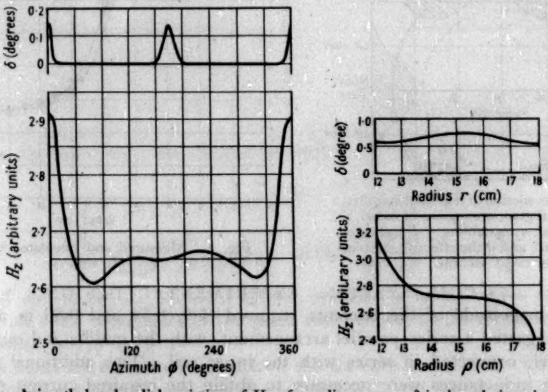


Fig. 7. Azimuthal and radial variation of H_z, after adjustment. δ is the relative phase of H_z. (Plane $z=0$; $r\equiv\rho$.)

It was found that the deviations of the field from the correct shape were just, but only just, within tolerable limits at 50 c/s. There were two undesirable effects. The first was due to out of phase fields caused by eddy currents in the connecting leads and probably also in the rather large diameter copper tubing used to make the main coils. The second was a rather large radial variation of the phase of the axial magnetic field, which could only partially be corrected by adjustment of the trimmer inductances. The net effect of these deviations is to limit the stable region at injection and during the first few revolutions of the electrons.

The field configuration produced by the pulse of current from the condenser discharge was not measured in detail. It was plain from the 50 c/s measurements that the eddy currents set up by this rapidly varying field would cause out of phase fields which would be far too large to allow injection into stable orbits to take place. The absence of an output from the machine under these conditions is therefore fully explained.

§7. CONCLUSIONS

The apparatus in its present form produces a sufficiently homogeneous field to permit operation at 50 c/s and low gas pressures in the doughnut. Modifications to the design of the coil system could undoubtedly be made which would considerably reduce the out of phase fields and the phase variations of the pulsed field. Operation at higher gas pressures would then be possible.

ACKNOWLEDGMENTS

The above work was carried out under the general supervision of Professor Sir George Thomson. The authors wish to thank him for his assistance and encouragement during the experiments, and the Ministry of Supply for a grant which made possible the construction of the betatron.

REFERENCES

Blewett, J. P., 1947, *J. Appl. Phys.*, **18**, 968.
Goward, F. K., 1948, *Proc. Phys. Soc.*, **61**, 284.

Magnetic Electron Microscope Projector Lenses

By G. LIEBMANN

Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berks.

MS. received 16*th July* 1951

ABSTRACT. The results of an earlier paper are extended to the special case of the electron microscope projector and similar type lenses. It is shown that the focal properties of symmetrical projector lenses with different ratios of gap width S to lens bore D can be represented by a single curve. This is then applied to the discussion of various cases of practical interest. The calculated results are found to be in excellent agreement with measurements by Ruska. The question of image distortion as a function of design parameters and lens excitation is then discussed; it is shown that all the types of lenses investigated can be operated at a point where they are free from 'radial distortion'. A consideration of scaling operations shows that there is little difference between the optimum performance of lenses of different geometry, but equal focal length. Certain results of use in practical lens design are combined in a design chart from which the relevant design or operational parameters can be read directly. In conclusion, design rules are given for the optimum design of projector lenses covering a wide range of magnifications.

§ 1. INTRODUCTION

In a preceding paper (Liebmann and Grad 1951, referred to below as I), the imaging properties of a series of symmetrical magnetic electron lenses were discussed. A subsequent short paper (Liebmann 1951) considered the particular case of the optimum electron microscope objective lens. Another special class of lenses, electron microscope projector lenses, will be investigated in this paper. Projector lenses are in a class of their own because they are often operated near their minimum focal length $(f_1)_{\min}$*. The question arises of designing lenses of shortest possible focal length, or shortest focal length for a given number of ampere turns, or of lenses with the smallest distortion for a given size of field and prescribed range of magnifications.

§ 2. THE GENERAL *f*-CURVE

Curves of f_1/R as function of the excitation parameter

$$k^2 = 0{\cdot}022\, H_0{}^2R^2/V_{\mathrm{r}} = \beta(NI)^2/V_{\mathrm{r}} \qquad \ldots\ldots(1)$$

were given in fig. 18 of I for four values of the relative pole piece separation S/D. Each of these curves shows a characteristic minimum focal length $(f/R)_{\min}$ at a particular value $k^2{}_{\min}$ of the excitation parameter.

If now the focal length for each of these curves is plotted relatively, as the ratio $f/f_{\min}$ of the focal length to the minimum focal length, as a function of the relative excitation parameter $p^2 = k^2/k^2{}_{\min} = (NI)^2/(NI)^2{}_{\min} = (V_{\mathrm{r}})_{\min}/V_{\mathrm{r}}$, it is found that all four cases are represented by one *single* curve, the 'general

---

* The notation used here is the same as that used in I.

† The focal length f_1 is the focal length of the lens when its whole length of field is used in the image formation, as shown in fig. 11 of I; the suffix 1 will be omitted in the following. The parameter β was plotted in fig. 2 of I.

f-curve' (fig. 1). The maximum deviations from the general *f*-curve are marked in fig. 1 by short bars above and below the mean curve, but these deviations are not of a systematic nature and merely express the inaccuracy due to interpolation of the original graphs. From this 'general *f*-curve' can be obtained the focusing properties of projector lenses of any arbitrary value of relative pole piece gap S/D, as $k^2_{\min}$ and $(f/R)_{\min}$ are known as functions of S/D.

These two functions, $(f/R)_{\min}$ and $k^2_{\min}$, are plotted in figs. 2 and 3. On investigating the magnetic field along the axis for large relative pole piece separations S/D, it is seen that the imaging properties of a lens with $S/D \gg 1$ approach closely those of a parallel uniform field of length S. As the focus for a ray incident parallel to the optical axis cannot lie far inside the lens field for the condition of minimum focal length, it can be assumed that $(f/R)_{\min} \simeq S/D$ for $S/D \gg 1$; in this case the 'weak lens formula' can be applied to give $f/R \simeq k^{-2}(S/D)^{-1}$, or $k^2_{\min} \simeq (S/D)^{-2}$ for $S/D \gg 1$. These two relations for $(f/R)_{\min}$ and $k^2_{\min}$, for large values of pole separation, help in extrapolating the curves of figs. 2 and 3 beyond $S/D = 2$, the last fully computed points.* It is seen that the conditions for $S/D \gg 1$ are already closely approached for $S/D \simeq 2$.

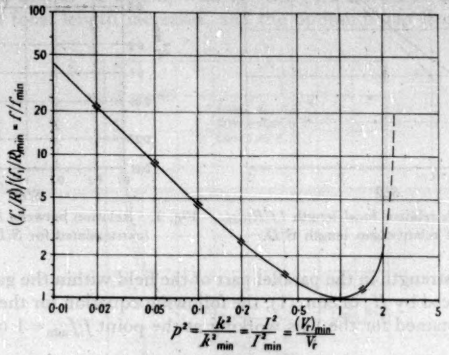


Fig. 1. General *f*-curve.

The paraxial focusing properties of the usual types of projector lenses can be completely derived from this set of figs. 1 to 3, except for the influence of iron saturation; several typical applications will be discussed in the following sections, and a design chart based partly on these curves will be given in fig. 12. As iron saturation leads essentially to an apparent increase in the relative pole separation S/D for very high excitation values, particularly for small values of S, its effect can often be estimated from the properties of the iron circuit used.

The magnification M for a screen distance b from the centre of the lens can usually be taken, with sufficiently high accuracy, as $M = (b/f) - 1$.

§ 3. THE MINIMUM POSSIBLE FOCAL LENGTH

As a first application, consider the ***minimum possible*** focal length of a symmetrical projector lens. If an unlimited maximum field strength H_0 could be used, the shortest possible focal length would be achieved with the shortest relative gap S/D, as is shown by fig. 2. But the limit set by iron saturation to the maximum field strength which can be employed leads to the existence of an optimum lens diameter $D_0 = 2R_0$ and the requirement of a long relative gap S/D, as was found previously for objective lenses of lowest spherical aberration (Liebmann 1951).

---

* Later computed values for $S/D=4$ of $(f/R)_{\min}$ and $k^2_{\min}$ agreed very well with the extrapolated values.

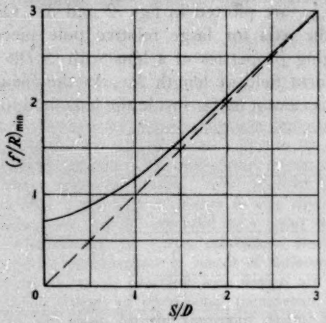


Fig. 2. Minimum relative focal length $(f/R)_{\min}$ as function of relative gap length S/D.

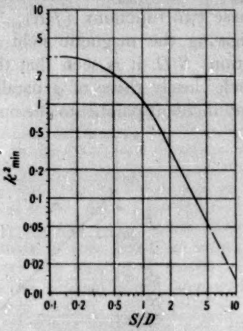


Fig. 3. Relation between $k_{\min}^2$ and S/D (extrapolated for $S/D>2$).

If the field strength in the parallel part of the field within the gap is called H_p, and H_0 is replaced by H_p in eqn. (1), the following equation for the optimum lens radius R_0 is obtained for the lens working at the point $f/f_{\min}=1$ of fig. 1, where $k^2 = k^2_{\min}$:

$$R_0 = B V_r^{1/2} H_p^{-1}\,[\mathrm{cm}]. \qquad \ldots\ldots(2)$$

In this equation, B is a known factor depending only on S/D:

$$B = \left(\frac{k^2_{\min}}{0{\cdot}022}\right)^{1/2} \left(\frac{H_0}{H_p}\right)^{-1}{}^{*} \qquad \ldots\ldots(3)$$

The minimum obtainable focal length, in centimetres, is then obviously $f_{\min} = (f/R)_{\min} R_0$, or from eqn. (2):

$$f_{\min} = B (f/R)_{\min} V_r^{1/2} H_p^{-1}\,[\mathrm{cm}] \qquad \ldots\ldots(4)$$

Eqn. (4) shows that $f_{\min}$, in absolute units, is directly proportional to the square root of the (relativistically corrected) accelerating voltage $V_r\,[V_r = V(1+10^{-6}V)]$ and inversely proportional to the field strength H_p measured in gauss. Further, it is directly proportional to the product $B(f/R)_{\min}$, which is a function of S/D only (see figs. 2 and 3). The geometrical factors $B = R_0 H_p V_r^{-1/2}$ and $B(f/R)_{\min} = f_{\min} H_p V_r^{-1/2}$ are plotted in fig. 4 as functions of S/D. It is seen that for given values of V_r and H_p the minimum possible focal length decreases with an increase in relative pole separation S/D; however, beyond $S/D \simeq 2$, this decrease becomes so small that $f_{\min}$ is practically independent of the relative pole separation, and depends only on V_r and H_p.

It is found that for $S/D>2$, the gap length S in absolute units, e.g. in centimetres, remains very nearly constant for variations of lens diameter if V_r

---

* H_0/H_p was plotted as function of S/D in fig. 7 of I.

and H_p are kept constant, and the lens geometry is adjusted to give the shortest possible focal length. This follows from the fact that in this region $B(f/R)_{\min} \simeq$ constant, or $B \propto R$. Further here $H_0/H_p \simeq 1$, and $B \propto k_{\min}$; also $k_{\min} \propto (S/D)^{-1}$, from fig. 3. Hence $S \simeq$ constant for $f_{\min}$ = constant, although S/D may be varied, i.e. R may be varied.

This can also be verified by working out several cases from the two curves given in fig. 4. Curve 1 gives $f_{\min}$ directly, if the value read off is divided by $H_p V_r^{-1/2}$, whereas curve 2 gives in the same way the corresponding value of R_0 which has to be employed to achieve this minimum focal length. As numerical example, take $V = 90$ kv ($V_r = 9{\cdot}8 \times 10^4$) and $H_p = 20\,000$ gauss; then $f_{\min} = 0{\cdot}118$ cm and $S = 0{\cdot}25$ cm, with $D = 2R < 0{\cdot}1$ cm. For larger lens bore, the minimum focal length increases, and the optimum gap length increases.

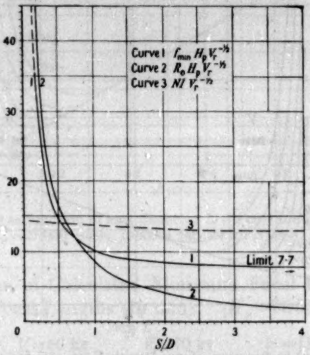


Fig. 4. Minimum possible focal length (curve 1) and corresponding optimum lens radius (curve 2) as function of relative gap width S/D, for given magnetic field strength and given accelerating voltage. Curve 3 gives required number of (nominal) ampere turns.

For $V = 60$ kv ($V_r = 6{\cdot}3_5 \times 10^4$) and $H_p = 26\,000$ gauss, the calculated values are $f_{\min} = 0{\cdot}75$ mm and $S = 1{\cdot}50$ mm. The minimum focal length measured by Ruska (1944) for these conditions was 0·75 mm, and the optimum gap length 1·25 mm (this slightly smaller experimental value is due to the neglect of iron saturation in the calculated value).

The nominal number of ampere turns NI needed to excite the lenses to give the minimum focal length can be obtained from fig. 3 and the graph of β as function of S/D given in I. The term $NI\, V_r^{-1/2}$ is represented by the broken curve in fig. 4; the (nominal) number of ampere turns required decreases slightly with increasing pole separation. As a numerical example, take again $V = 90$ kv and $H_p = 20000$ gauss. Then for large values of pole piece separation, $NI = 4080$ ampere turns. In practice, the m.m.f. drop in the iron circuit, particularly if part of the iron circuit should become saturated, will necessitate a somewhat increased value of ampere turns; for lenses with short pole-piece gaps, where the reluctance of the air gap is not high compared with the reluctance of the iron circuit, this increase will be relatively greater than for lenses with long gaps.

§ 4. VARIATION OF FOCAL LENGTH WITH CHANGE OF LENS DIAMETER OR CHANGE OF GAP WIDTH

Two further applications of the 'general *f*-curve' are given in the two sets of graphs, figs. 5 and 6. The focal length of projector lenses of fixed diameter $D=0{\cdot}2$ cm is plotted in fig. 5 for several accelerating voltages as a function of gap width S. In fig. 6, the focal length is plotted, for the fixed relative gap width $S/D=1$, as a function of the lens diameter D. In figs. 5 and 6 a field strength $H_p=(H_p)_{\max}=25\,000$ gauss was assumed. Similar curves are obtained for other values of H_p; for example, a decrease in H_p would move the curves in both figures to the right, the amount of shift being such that $(H_pV_r^{-1/2})$ remains a constant for each of the curves shown.

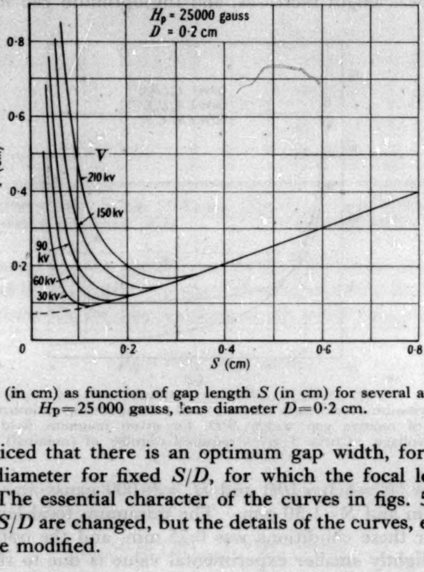


Fig. 5. **Focal length (in cm) as function of gap length S (in cm) for several accelerating voltages. $H_p=25\,000$ gauss, lens diameter $D=0{\cdot}2$ cm.**

It will be noticed that there is an optimum gap width, for fixed diameter, or an optimum diameter for fixed S/D, for which the focal length attains its smallest value. The essential character of the curves in figs. 5 and 6 remains unaltered if R or S/D are changed, but the details of the curves, e.g. the positions of the minima, are modified.

§ 5. COMPARISON WITH MEASUREMENTS

The extensive measurements of Ruska (1944) are in good agreement with our calculations; e.g. table 1 gives a comparison of data calculated from our general *f*-curve, and of some of Ruska's experimental data, interpolated from fig. 7 of his paper. A value of $(H_p)_{\max}=26\,000$ gauss* was assumed.

The influence of iron saturation is relatively unimportant for either low voltages or large gap widths, and agreement here is excellent. At high voltages and small gap widths, the influence of iron saturation becomes more marked, and agreement is less good. As mentioned before, the result of iron saturation is to make the gap appear greater than it is in fact*, and if this is taken into account, agreement between our calculated values and Ruska's measurements becomes very satisfactory even at the high voltages and small gap widths. (For the higher voltages, the number of ampere turns used in Ruska's experiments was more than twice the number for which saturation had become noticeable.)

---

* Fig. 5 of Ruska's paper gives a value of $(H_0)_{\max}=25\,000$ gauss for $S/D=1$, which would require $(H_p)_{\max}=30\,000$ gauss. An interpretation of the experimental data in the light of our computations would give values of $(H_p)_{\max}$ in the range of 23 000 to 27 000 gauss for the different sets of Ruska's curves.

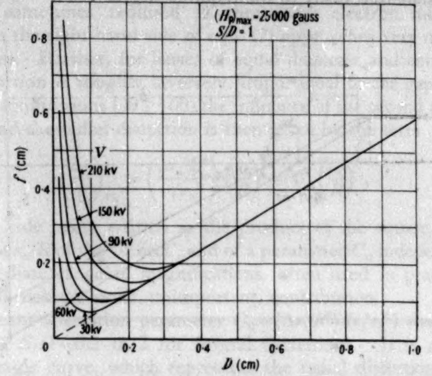


Fig. 6. Focal length (in cm) as function of lens diameter D (in cm) for several accelerating voltages. $H_\mathrm{p}=25\,000$ gauss, relative gap width $S/D=1$.

Table 1. Comparison of Calculated Minimum Focal Lengths and Ruska's Measured Minimum Focal Lengths (in mm). $H_\mathrm{p}=26\,000$ gauss, $D=0{\cdot}6$ mm.

	V=40 kv		V=60 kv		V=90 kv		V=150 kv		V=210 kv	
S/D	calc.	meas.	calc.	meas.	calc.	meas.	calc.	meas.	calc.	meas.
1	0·92	0·82	1·30	1·06	1·92	1·30	3·33	—	5·0	—
2	0·63	0·66	0·74	0·79	0·95	0·93	1·47	1·25	2·02	1·65
3	0·91	0·90	0·91	0·92	0·93	1·0	1·16	1·18	1·51	1·5
4	1·21	1·24	1·21	1·27	1·21	1·30	1·24	(1·4)†	1·43	(1·6)†

† Extrapolated.

The focal length curves calculated by van Ments and le Poole (1947) on the basis of measured field distributions also agree closely with our results, except for small gap widths and weak excitations, and for great gap widths and very strong excitations, where small discrepancies appear.

§ 6. CHROMATIC CHANGE OF MAGNIFICATION

It is obvious from an inspection of fig. 1 that, at the point $k^2/k^2_{\min}=1$, all projector lenses are free from a chromatic change of magnification. However, even when not operated at this excitation value, the chromatic aberration constant C_c, which gives rise to the chromatic change of magnification according to the relation

$$\Delta M/M=(C_c/f)(2\Delta H/H-\Delta V/V) \quad \ldots\ldots(5)$$

---

* This effect amounts to an increase of gap width by 10% to 40% for $S/D=1$, depending on the operating conditions, and less for larger values of S/D.

is always smaller than f (its limiting value for very weak excitation), tending to zero as $k^2/k^2_{\min}=1$ is approached. The chromatic change of magnification in the projector lenses is therefore quite negligible in modern electron microscopes as these use a high degree of accelerating voltage stabilization to ensure satisfactory operation of the objective; the lens current of the projector lenses can always be stabilized easily to the required degree.

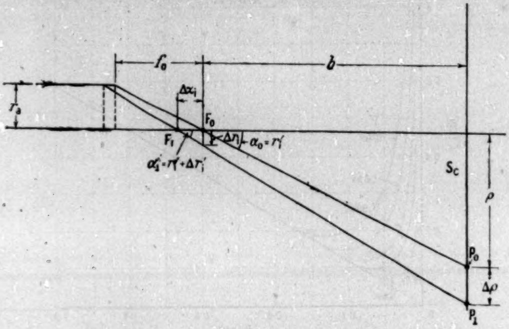


Fig. 7. Ray diagram illustrating the origin of the radial distortion $\Delta\rho/\rho$.

The same can be said of the chromatic change in image rotation Ψ, although its effect may be greater than that of the chromatic change of magnification.

The 'chromatic' displacement δ_c of an image point at a distance ρ from the axis is

$$\delta_c=\tfrac{1}{2}\rho(2\Delta H/H-\Delta V/V)\Psi, \qquad \ldots\ldots(6)$$

with $\Psi=(k/k_{\min})\Psi_0$, where Ψ_0 is the rotation at $k^2/k^2_{\min}=1$. The values of Ψ_0 (in radians) for the four lenses $S/D=0{\cdot}2$, $0{\cdot}6$, 1, 2 are 4·72, 3·61, 2·76, 1·41 respectively.

§7. RADIAL DISTORTION

Electron microscope projector lenses suffer from distortion of the image field, and a determination of the design parameters which lead to low distortion values is of importance. The two components of distortion, which may be called 'radial distortion' and 'spiral distortion'*, will be dealt with separately.

Consider in the schematic ray diagram, fig. 7, a ray incident parallel to the optical axis, at a distance r_a from it, all dimensions being measured relatively, for instance in terms of unit R, the lens radius. If $r_a/R\ll 1$, the ray will pass through the paraxial focus F_0 and intersect the image screen Sc, placed at the distance $b=(M+1)f$ from F_0, in the point P_0 at the distance ρ from the axis. For a larger off-axis distance r_a of the incident ray, it will pass through the focal point F_1, which is shifted by $\Delta z_1=C_s\alpha^2=C_s r_a^2/f^2$ from F_0, and intersect the screen at P_1 distant from the axis $\rho+\Delta\rho$. Its angle of intersection with the axis will now be $\alpha_1=r_1'+\Delta r_1'$, where $\Delta r_1'$ is the deviation of the tangent of the 'first order ray' from the tangent $r_1'=\alpha$ of the paraxial ray. We then have

---

* Spiral distortion was called 'circumferential' distortion by Hillier (1946) in his study of projector lens distortion.

$r_1' = \rho/b$, and $r_1' + \Delta r_1' = (1+\Delta\rho/\rho)/b(1+C_s r_1'^2/b)$ whence we obtain for the radial distortion $(\Delta\rho/\rho)$

$$\frac{\Delta\rho}{\rho} \simeq \frac{\Delta r_1'}{r_1'} + \frac{C_s}{b} r_1'^2. \qquad \dots\dots(7)$$

Reference to figs. 20 and 23 of I shows that for very low magnifications, $M \to 1$ $(k^2 \ll 1)$, sometimes required in three-stage electron microscopes, the second term on the right-hand side of eqn. (7) approaches very nearly the value of the first term. Further, for lenses of equal diameter and equal focal length the radial distortion is roughly inversely proportional to the gap width S/D.

For high magnifications $(M \gtrsim 100)$ the influence of the second term in eqn. (7) is negligible, and the radial distortion is then given by the term

$$\left(\frac{\Delta\rho}{\rho}\right)_{r_a} = \left(\frac{\Delta r_1' R^2}{r_1' r_a^2}\right)\left(\frac{r_a}{R}\right)^2 = C_d\left(\frac{r_a}{R}\right)^2, \qquad \dots\dots(8)$$

the right-hand side being written as the product of the square of the relative off-axis distance r_a/R of the 'object', and of a parameter C_d independent of r_a/R*. In the intermediate range of magnifications, often used in practice, the term neglected in (8) makes a small, unimportant, contribution.

The significant aberration parameter $C_d = (\Delta r_1' R^2)/(r_1' r_a^2)$ was computed for the four lenses $S/D = 0{\cdot}2$ to 2 for several excitations. It is not possible to abstract one single curve, which represents the radial distortion parameter C_d as a function of the excitation ratio $k^2/k^2_{\min}$, but it facilitates a comparison of the different lenses and interpolation of results for cases other than those computed if the ratios C_d/C_{d_0} are plotted as a function of $k^2/k^2_{\min}$ for the four lenses (fig. 8); here C_{d_0} is the value of the radial distortion parameter for each lens for very small excitations, $k^2 \to 0$. To convert the curves, fig. 8, into absolute values, the constant C_{d_0} is plotted in fig. 9 as a function of S/D.

It is seen from fig. 8 that C_d/C_{d_0} remains nearly constant for weak and moderately strong relative excitations $k^2/k^2_{\min}$, i.e. the relative distortion $\Delta\rho/\rho$, referred back to the 'object' with off-axis distance r_a/R (eqn. (7)), remains nearly constant. When measured in the final image for a fixed size of object field, the distortion is therefore approximately inversely proportional to the magnification of the projector lens. For a relative excitation value $k^2/k^2_{\min} \simeq 0{\cdot}5$, the relative distortion C_d/C_{d_0} reaches a maximum, and then drops away fairly steeply. Near $k^2/k^2_{\min} \simeq 1$, the distortion constant C_d changes its sign, i.e. 'pincushion' distortion changes to 'barrel' distortion. For each lens there is a certain excitation value, near the value $k^2_{\min}$, for which the radial distortion vanishes†. (The contribution of the term neglected in eqn. (8) is negligible for these relatively high excitation values and consequently high magnifications.) For $S/D \simeq 1{\cdot}8$, this excitation value coincides exactly with the value $k^2_{\min}$, and for greater gap widths, the excitation for operation of the lens without radial distortion is slightly smaller than that needed to reach the minimum focal length. The slope of the curves (fig. 8) near their intersection points with the line $C_d = 0$ becomes smaller for lenses with greater gap widths, making the operation of a projector lens near this point less critical as far as reduced radial

.B
55
52

---

* C_d is connected with the distortion parameter $S-T$, used in a previous paper on electrostatic lenses (Liebmann 1949), by the relation $C_d = (R/f)^2(S-T)$.

† A similar distortion free operating range had been found experimentally in electrostatic projector lenses by Rang (1948–49).

distortion is concerned. Hillier (1946) described a combination of two projector lenses which formed an image free from radial distortion, for a certain operating condition. From fig. 8 we see that *any* single lens can be operated to give this effect, provided that iron saturation does not limit the excitation to a value lower than that required for this particular lens. The appropriate design data for the operation of the lens in the distortion-free mode can be derived from the data given earlier in this paper. For instance, for the lens $S/D=1{\cdot}8$, $R=5{\cdot}6\,V_r^{1/2}H_p^{-1}$ [cm] is found; e.g. $D=1{\cdot}76$ mm for $V=90$ kv and $H_p=20\,000$ gauss, requiring 4350 ampere turns (nominal value). If a lens were to be operated at the distortion-free point with a smaller number of ampere turns, its relative gap width would have to be rather greater.

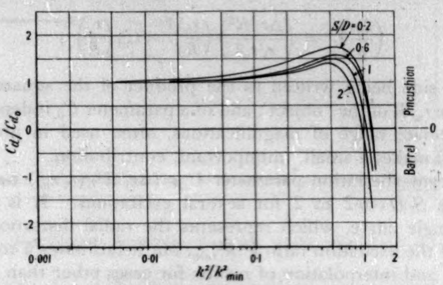


Fig. 8. Relative radial distortion coefficient C_d/C_{d_0} as function of $k^2/k^2_{\min}$ for several values of relative gap width S/D.

The data of figs. 9 and 10 can be used to predict the radial distortion of a given projector lens for prescribed working conditions, or to lay down maximum distortion values in a design specification. For instance, if it were specified that the distortion in an image point corresponding to $r_a/R=0{\cdot}25$ should be $\Delta\rho/\rho \leqslant 0{\cdot}02$ for *all* possible operating conditions ($k^2<k^2_{\min}$), it is found that $(C_d)_{\max} \leqslant 0{\cdot}32$, $C_{d_0} \leqslant 0{\cdot}22$, and $S/D \geqslant 1{\cdot}2$. For $\Delta\rho/\rho \leqslant 0{\cdot}01$, it is found that $S/D \geqslant 1{\cdot}9$.

§ 8. SPIRAL DISTORTION

The image rotation, which is characteristic of magnetic lenses, increases by a small amount $\Delta\Psi$ for an extra-paraxial ray, thus giving rise to the anisotropic first-order image defects (Glaser 1933). Under the imaging conditions of the electron microscope projector lens, as shown in fig. 7, $\Delta\Psi$ is proportional to $(r_a/R)^2$: $\Delta\Psi=C_{sp}(r_a/R)^2$, where C_{sp}, the 'spiral distortion constant', depends on the lens geometry and the excitation.

The lateral displacement δ in the image plane Sc (fig. 7) is then $\rho\Delta\Psi$, or $\delta/R=MC_{sp}(r_a/R)^3$. A straight line in the image plane passing through the axis is thus turned into a cubic parabola (fig. 10). If the displacement δ is referred to the image point distance from the axis, ρ, then obviously $\delta/\rho=\Delta\Psi$, and δ/ρ is described by eqn. (9),

$$(\delta/\rho)_{r_a}=C_{sp}(r_a/R)^2 \qquad \ldots\ldots(9)$$

which is analogous to eqn. (8).

The spiral distortion constant C_{sp} was computed for the four lenses $S/D=0{\cdot}2$ to $S/D=2$ for several excitations. If the result is expressed in relative units, by plotting C_{sp}/C_{sp_0} as a function of $(k/k_{min})^2$, where C_{sp_0} is the value of the spiral distortion constant for $(k/k_{min})^2=1$, one *single* curve is obtained, which is a straight line on a double logarithmic scale (fig. 11). In contrast to the radial distortion, spiral distortion continues to increase as the excitation is increased, but it is again seen that the longer lenses are less subject to this aberration than are the shorter lenses. This is shown more clearly by the curve marked C_{sp_0} in fig. 9, which gives the C_{sp_0} values as a function of S/D. By combining this curve with the graph (fig. 11) the spiral distortion constant

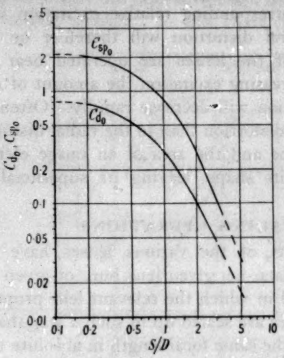


Fig. 9. Radial distortion coefficient C_{d_0} and spiral distortion coefficient C_{sp_0} as function of relative gap width S/D.

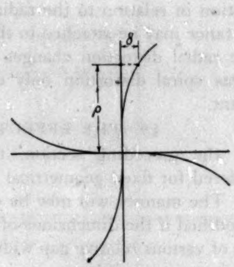


Fig. 10. Diagram illustrating spiral distortion δ/ρ.

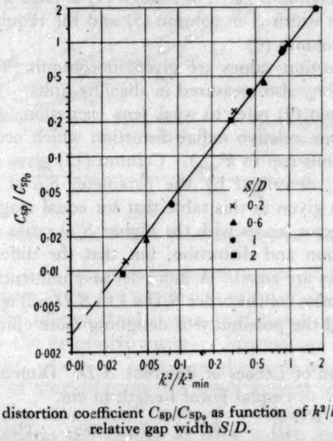


Fig. 11. Relative spiral distortion coefficient C_{sp}/C_{sp_0} as function of k^2/k^2_{min} for several values of relative gap width S/D.

can be obtained for any relative gap width and excitation likely to be met in practice.

As an example, consider the application of these results to the two cases discussed at the end of the preceding section. When $S/D=1{\cdot}2$, $C_{\mathrm{sp}_0}=0{\cdot}70$, and when $S/D=1{\cdot}9$, $C_{\mathrm{sp}_0}=0{\cdot}31$. $(\delta/\rho)_{\max}$ has the same value as the $\Delta\rho/\rho$ specified in the example for $r_a/R=0{\cdot}25$, if $C_{\mathrm{sp}}/C_{\mathrm{sp}_0}=0{\cdot}46$ in the first case and $C_{\mathrm{sp}}/C_{\mathrm{sp}_0}=0{\cdot}52$ in the second case, the corresponding relative excitation values being $k^2/k^2_{\min}=0{\cdot}56$ and $0{\cdot}60$. The spiral distortion will therefore be more pronounced than the radial distortion if the lenses are operated near their minimum focal length; however, for decreasing excitation the amount of spiral distortion in relation to the radial distortion will decrease rapidly. Often, less importance may be attached to the spiral distortion than to the radial distortion, as the radial distortion changes the shape and the area of an image element, whereas spiral distortion only changes its shape, leaving its superficial area constant.

§ 9. THE EFFECT OF SCALING OPERATIONS

In the preceding section, the merits of the various lenses have been considered for fixed geometrical design data, i.e. given lens bore or given ratio S/D. The manner will now be discussed in which the relevant lens properties are modified if the dimensions of the lenses are scaled up in such a way that the lenses of various relative gap widths give the same focal length in absolute units, i.e. the same magnification.

According to fig. 1, the lens dimensions have to be scaled in such a ratio that $f=(f/R)_{\min}\times R\times$ constant is the same for all values of S/D, or $R\propto(f/R)_{\min}^{-1}$. The result of this operation is given in column (4) of table 2 which also gives the resulting relative gap width S in column (5) and the required relative number of ampere turns in column (6).

The relative distortion values are given in columns (9), (10) and (12), all for the same value of r_a, also measured in absolute units. The radial distortion values given in column (9) refer to weak lens excitation, whereas column (10) refers to the maximum relative radial distortion which occurs over the whole range of lens excitations (up to $k^2_{\min}$). Column (12) gives the relative value of the spiral distortion, described by the parameter δ/ρ at $(k/k_{\min})^2=1$. It is seen from the results given in this table that for equal magnification and equal size of field the projector lenses with the higher S/D ratios are slightly superior in respect of excitation and distortion, but that the differences between the various types of lenses are small. A more decisive influence in favour of using relatively large S/D ratios (of the order $S/D=1$ to $S/D=2$) is the easier avoidance of iron saturation and the possibility of designing more efficient iron circuits.

Table 2. Comparison of Lenses of different S/D. Dimensions scaled to give equal Focal Length in cm.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
S/D	$(f/R)_{\min}$	$k^2/k^2_{\min}$	R_{rel}	S_{rel}	NI_{rel}	$(r_a/R)^2_{\mathrm{rel}}$	C_{d_0}	$\left(\frac{\Delta\rho}{\rho}\right)_{0\,\mathrm{rel}}$	$\left(\frac{\Delta\rho}{\rho}\right)_{0\,\max\,\mathrm{rel}}$	C_{sp_0}	$\left(\frac{\delta}{\rho}\right)_{0\,\mathrm{rel}}$
0·2	0·75	3·1	1·58	0·32	1·05	0·40	0·660	0·91	1·06	2·00	0·96
0·6	0·91	1·95	1·30	0·78	1·01	0·60	0·475	0·96	1·02	1·36	0·98
1	1·18	1·15	1·00	1·00	1·00	1·00	0·295	1·00	1·00	0·84	1·00
2	2·07	0·34	0·57	1·14	0·94	3·08	0·098	1·02	0·96	0·28	1·02

For columns (4)–(7) $f=$const., $r_a=$const.

§10. A DESIGN CHART FOR PROJECTOR LENSES

In the foregoing, the general properties of projector lenses have been discussed, and a few numerical examples have been given to illustrate the application of the various graphs to the study or to the design of projector lenses. Some of this information will now be put into the form of a design chart, with the help of which the relevant design questions of symmetrical projector lenses can be solved quickly. In this chart (fig. 12) the relative focal length f/D is plotted as a function of S/D for several constant values of the lens excitation parameter $NI\,V_r^{-1/2}$, where NI is the nominal number of ampere turns*. The numerals written by the side of these curves (full lines) denote the value of $NI\,V_r^{-1/2}$ applicable to the respective curves. Moving up vertically, for the appropriate value of S/D, shows at once how the focal length changes for a change in the number of ampere turns (for fixed voltage), or for a change of accelerating voltage (for fixed coil current). Moving along the lines $NI\,V_r^{-1/2} = \text{constant}$ indicates the change of focal length due to a change in lens geometry.

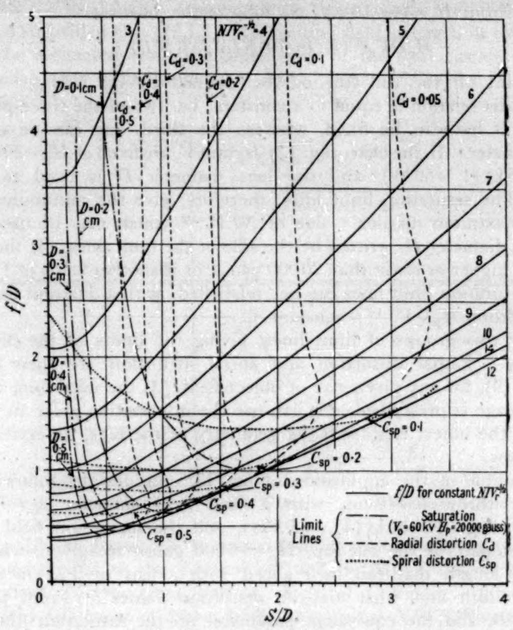


Fig. 12. Design chart for electron microscope projector lenses.

The chart also carries certain 'limit curves', which show over which range of lens excitations or for which minimum relative gap widths the lens may be used if certain limits of saturation field strength $(H_p)_{\max}$ (chain line) are prescribed, or if certain maximum values of radial distortion (short dashes) or spiral distortion (dots) are permitted over the whole working range of the lens. In each case, the part of the $NI\,V_r^{-1/2}$ curves to the right (or above in the case of spiral distortion) is permitted; if the $NI\,V_r^{-1/2}$ curves are followed down beyond the limit lines, the prescribed maximum distortion values are exceeded, or iron saturation becomes predominant.

---

* The nominal value of NI has to be increased in practice by the amount lost in the iron circuit. This may be of the order of 10% and more when the saturation flux is approached.

The limit lines referring to iron saturation (chain line) require further discussion. From $H_p=(4\pi/10)(NI/S)$ follows the relation

$$H_pDV_r^{-1/2}=(4\pi/10)NI\,V_r^{-1/2}(S/D)^{-1}.$$

For each point on the full line curves, defined by its $NI\,V_r^{-1/2}$ and S/D coordinates, this relation is equal to a constant, i.e. two of the three parameters H_p, V_r and D have to be fixed, whereas the third one can be used as a variable parameter. In the chart (fig. 12) H_p and V_r are fixed at $H_p=20\,000$ gauss and $V_r=63{\cdot}5$ kv ($V=60$ kv), and the lens diameter D is used as variable parameter. The saturation limit lines, therefore, give the minimum value of S/D or the maximum nominal value of $NI\,V_r^{-1/2}$ which may be used for the values of lens diameter D, written by the side of the limit lines. If the limiting value of H_p is higher or lower than 20 000 gauss, or if another value of V_r is used, the present saturation limit lines can be 'relabelled' with a different diameter D from the condition $H_pD\,V_r^{-1/2}=\text{constant}$.

The other two groups of limit lines, giving the limits of the distortions, apply directly. Radial distortion and spiral distortion are then given by eqns. (8) and (9), for any given size of object field. If the maximum distortion in the final image is prescribed, one can use the information given in fig. 12 to work back to the object size from the geometry of the lens, its excitation and the image throw.

As an example of the application of fig. 12, consider the following case: If the image throw $b=40$ cm, with a final field diameter $2\rho=7$ cm, the accelerating voltage $V=60$ kv ($V_r=63{\cdot}5$ kv), and the maximum field strength in the parallel part of the air gap, $H_p=16\,000$ gauss are given, what is the shortest focal length that can be realized with a lens of 0·25 cm bore and 0·20 cm gap width and what are the distortion values? From the given data $S/D=0{\cdot}80$, and the equivalent parameter on the saturation limit line is $(16\,000\times0{\cdot}25)/20\,000=0{\cdot}20$ cm. This limit line intersects the vertical at $S/D=0{\cdot}80$ at approximately $NI\,V_r^{-1/2}\simeq9{\cdot}7$. This gives a (nominal) maximum number of ampere turns $NI=2450$. The f/D value read off at this working point is $f/D=0{\cdot}69$, or $f=0{\cdot}173$ cm; hence the maximum magnification is $M=(b/f)-1=230$. The object field radius is then $r_a/R=2\rho/MD=0{\cdot}122$; from $C_d\simeq0{\cdot}53$ and $C_{sp}=0{\cdot}40$ read off in fig. 12 at this working point, $\Delta\rho/\rho=0{\cdot}8\%$ and $\Delta\delta/\rho=0{\cdot}6\%$ at the edge of the viewing screen.

Now consider, in this example, the increase in distortion if the magnification is reduced in the ratio 5 : 1. The new value of f/D is 3·45, with $NI\,V_r^{-1/2}\simeq3{\cdot}7$ and $NI\simeq930$, with $C_d\simeq0{\cdot}41$ and $C_{sp}\simeq0{\cdot}03$. Hence, at the edge of the viewing screen $\Delta\rho/\rho=15\%$ and $\Delta\delta/\rho\simeq1\%$. This radial distortion value is rather worse than the value of 9% which could be obtained if the lens were redesigned in such a way that the minimum focal length is not prematurely limited by iron saturation. This could, for instance, be done by increasing the gap length to 0·30 cm ($S/D=1{\cdot}2$), and increasing the (nominal) number of ampere turns to 3000. If the increased number of ampere turns were not available, it would still be possible to obtain a slight improvement, e.g. by increasing the gap length to 0·28 cm and decreasing the lens diameter to 0·175 cm, and by keeping the number of ampere turns and the focal length constant at their previous values of 2450 and 0·173 cm. In this case the maximum distortion at the edge of the field for a range of magnification of 5 : 1 would be reduced slightly to 12·5%.

The data of the design chart, fig. 12, can be applied to predict quite accurately the focal length–lens excitation characteristic even if the iron becomes saturated (provided that $(H_p)_{max}$, the limiting field strength in the gap, is known or can be estimated from the dimensions of the iron circuit). This will be explained by the example given in fig. 13, where $S/D = 0{\cdot}4$, $D = 0{\cdot}508$ cm and $(H_p)_{max} = 12\,500$ gauss due to saturation of part of the iron circuit. These values correspond to the dimensions used by Mr. J. Rawes of the Research Department, Metropolitan-Vickers Electrical Company. From the data of fig. 12, the f/D curve is drawn with the nominal $NI\,V_r^{-1/2}$ values as abscissae, for the specified value of S/D (dashed curve). A $NI\,V_r^{-1/2}$ scale is then superimposed, contracted to take account of the loss of ampere turns in the iron circuit (without saturation), e.g. by 5% as in fig. 13. Using this modified scale, the point A is found on the dashed curve which corresponds to the excitation value for which $(H_p)_{max}$ would be reached in the gap, viz. $[NI\,V_r^{-1/2}]^* = (10/4\pi)D(S/D)V_r^{-1/2}(H_p)_{max}$, and the (dotted) 'saturation line' is drawn through point A, parallel to the x axis. The points B and C, corresponding to $0{\cdot}5[NI/V_r^{-1/2}]^*$ on the dashed line, and $1{\cdot}5[NI\,V_r^{-1/2}]^*$ on the dotted line are then noted. Finally, points B and C are connected by a curve (full line) which is tangential to the dashed f/D curve at B and the dotted saturation line C, thus rounding off the sharp corner at A. This final curve is the required practical f/D curve, accurate for weak excitation $NI/V_r^{-1/2} \leqslant 0{\cdot}5[NI/V_r^{-1/2}]^*$, and a good approximation for higher excitations. How good this approximation is, can be seen from Mr. Rawes's measured points, shown in fig. 13.

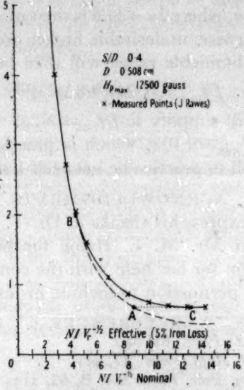


Fig. 13. Construction of practical f/D curve, in the presence of iron saturation.

§ 11. SOME RULES FOR OPTIMUM DESIGN

In conclusion, as the result of an analysis of the design chart (fig. 12) a few design rules can be drawn up for the case of as wide a range of magnifications as possible for a specified maximum distortion at the edge of the viewing field.

(i) *For widest range*, a number of (nominal) ampere turns should be provided which is at least 12 $V_r^{1/2}$, for $S/D \geqslant 1{\cdot}5$ and up to 20% higher for smaller S/D ratios. Since this nominal value does not take into account the drop of the m.m.f. in the iron circuit or saturation effects, the number of ampere turns used in practice has to be increased accordingly.

(ii) The minimum value of S/D should be chosen such that iron saturation is avoided, or is at least not very pronounced.

(iii) The maximum value of S/D should be such that at the low magnification end of the range $f/D \leqslant 1/4\alpha$, where $2\alpha = 2\rho/b$ is the ratio of viewing screen diameter to screen distance. Otherwise, undesirable higher order aberrations will appear.

(iv) The maximum obtainable range will then be given by

$$f/f_{\min} = (1/2\alpha)(9{\cdot}8\Delta\rho/\rho)^{1/2}.$$

For $\Delta\rho/\rho = 10{\cdot}2\%$, this will simplify to $f/f_{\min} = b/2\rho$. (These relations are based on the fact that $C_{d_e}(f/D)^2{}_{\min} = 0{\cdot}102$, which is practically independent of S/D.)

The best S/D ratio will in practice lie between 1 and 2.

ACKNOWLEDGMENTS

The writer wishes to express his thanks to Dr. T. E. Allibone for permission to publish this paper, to Mr. M. E. Haine for helpful discussions, and to Miss M. M. MacNaughton for her help with the computational work. He also thanks Mr. J. Rawes for permission to include his experimental data in fig. 13.

REFERENCES

Glaser, W., 1933, *Z. Phys.*, **81**, 649.

Hillier, J., 1946, *J. Appl. Phys.*, **17**, 411.

Liebmann, G., 1949, *Proc. Phys. Soc.* B, **62**, 213; 1951, *Ibid.*, **64**, 972.

Liebmann, G., and Grad, E. M., 1951, *Proc. Phys. Soc.* B, **64**, 956.

van Ments, M., and le Poole, J. B., 1947, *Appl. Sci. Res.*, **B.1**, 3.

Rang, O., 1948–49, *Optik*, **4**, 251.

Ruska, E., 1944, *Arch. Elektrotech.*, **36**, 431.

The Electrical and Luminous Characteristics of Short Air Sparks suitable for High Speed Photography

By W. G. STANDRING and J. S. T. LOOMS
National Physical Laboratory, Teddington, Middlesex

Communication from the National Physical Laboratory ; *MS. received* 12*th April* 1951, *and in amended form* 18*th September* 1951

ABSTRACT. The light emitted by various forms of spark gap is examined, and the factors affecting maximum value and variation with time are discussed. Measurements of the voltage drop across the spark help to explain the effects on light output of spark duration and form of spark gap. A rough spectral analysis is given for an open gap.

§ 1. INTRODUCTION

In high-speed photography much use is now made of short duration flashes from electric discharges. For photography by reflected light high-pressure gas discharge tubes are available, but are expensive and have a limited life. In many investigations shadow photography requiring a point source is used, or schlieren technique for which a short line source is sufficient. For these purposes a suitable and convenient light source is a short spark in air at atmospheric pressure. Electrodes to provide such sparks are easily set up and easily replaced.

To assist investigators using light sources of this type in designing equipment to produce a flash of the required intensity and duration, a study has been made of the relation between the light output and the electrical characteristics of the discharge, and also of the effects of the material and form of the electrodes. The duration considered is of the order of microseconds. The sparks examined were obtained by the oscillatory discharge of condensers, this giving the highest maximum light intensity for a given condenser charged to a given voltage. Particular attention was given to sparks carrying damped oscillatory currents of about 4 kiloamperes with an oscillation period of about 1 microsecond which were found satisfactory for use in the wind tunnels of the Aerodynamics Division of the National Physical Laboratory.

§ 2. FORMS OF SPARK GAP EXAMINED

The gap studied in most detail was a 1 mm gap between two electrodes, usually rods 6 mm diameter with hemispherical ends.

In another 'point source' examined the discharge took place in a constricting tube of insulating material of internal diameter 1 mm. One electrode entered one end of the tube. The other electrode in contact with the other end of the tube contained a coaxial hole of 1 mm diameter through which the light emerged. (Without the constricting tube, the discharge was usually not axial and less light emerged through the hole than was obtained from a 1 mm open gap.)

Observations were also made on a 'line source' obtained by a discharge between metal foil electrodes clamped between glass plates, viewed in a direction parallel to the plates.

§ 3. MEASURING TECHNIQUE

The total light output was measured with a portable integrating photoelectric flashmeter (Preston 1943) giving candle seconds on the standard photometric scale (light adapted eye). The variation of light intensity with time was measured by means of a photomultiplier, type RCA 931A, a four stage amplifier and a cathode-ray oscillograph. For voltage measurements the gap electrodes were connected by a concentric cable to a pair of oscillograph plates which would withstand a voltage of 10 kv. Connection to the gap electrodes was made as near as possible to the electrode tips and observations of the inductive volt drop between the tapping points were made by filling the gap with a wedge. Current measurements were obtained from oscillograms of the voltage across a 0·02 or 0·05 ohm shunt, made of bifilar strip, included in the discharge circuit. Power and resistance data were obtained by subtracting the inductive volt drop from the voltage observed and multiplying or dividing by the current.

§ 4. LIGHT OUTPUT

(i) *Total Light*

Flashmeter measurements with different electrode materials showed that the light output was roughly related to the boiling point of the metal used, mercury giving most light and copper or tungsten least. Photomultiplier measurements showed that in general the light from the more volatile metals reached a higher peak and decreased more slowly (see curves 3 and 4 in fig. 2 (*a*) for a comparison of copper and cadmium). Integration of the photomultiplier oscillograms gave relative agreement with the flashmeter measurements.

With a spark current of oscillation period 1·2 microsecond, and successive maxima 4, 3·3, and 2·8 kA, the light output from a 1 mm open gap for copper electrodes, for cadmium electrodes, and for one mercury and one cadmium electrode, was 0·1, 0·2 and 0·3 candle second. For constricted gaps 1·5 and 8 mm in length values of 0·6 and 0·9 candle second were obtained.

(ii) *Variation of Light with Time*

For a given gap the light–time characteristic depends on the current–time characteristic. The same light output was obtained by the discharge of a 0·1 μF condenser at 10 kv through a circuit inductance of 0·4 μH as by the discharge of a 0·033 μF condenser at 30 kv through a circuit inductance of 1·2 μH, giving the same current of 5 kA with an oscillation period of 1·3 μsec.

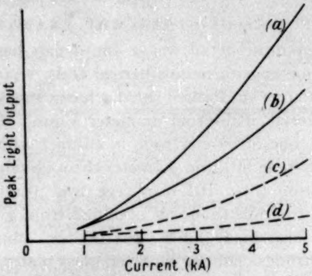


Fig. 1. Variation of peak light with current in 1 mm open gap : (*a*), (*c*) cadmium electrodes, (*b*), (*d*) copper electrodes; (*a*), (*b*) discharge oscillation period 1·3 μsec, (*c*), (*d*) discharge oscillation period 40 μsec.

For a damped sinusoidal current of given oscillation period, the maximum light intensity varies with the maximum current in the manner shown in fig. 1. For a given maximum current the light peak increases as the time of rise of current decreases over the range 12 to 0·3 microseconds. For shorter times it appears to decrease.

Fig. 2 (*a*) gives light–time characteristics for a gap current consisting of a damped sinusoid with successive peaks of 4, 3·3, and 2·8 kA, and with an oscillation period of either 1·2 (short flash) or 48 (long flash) microseconds. The light increases and decreases with the current, but the rate of decrease is limited and with oscillation periods of the microsecond order the light does not fall to zero at zero current. Curves 1 and 2, giving light characteristics of an open gap for a long period flash, show average results. With the long flash, individual observations are sometimes irregular in shape; with copper electrodes the highest light peak may occur in the first, second or third half cycle; with cadmium electrodes the first light peak is sometimes little more than the second or third. For a short period flash (curves 3 and 4) the first light peak is much the highest. With a constricted gap, for a short period flash (curves 5 and 6) the light output shows peaks of the same order for several half cycles. The line source mentioned in §2 gave a light characteristic intermediate between those of open and constricted gaps.

In fig. 3 is shown the light output from a 1 mm open gap carrying a current of amplitude 8 kA and oscillation period 0·23 microsecond; the light peak is of about the same magnitude as that for a current of 4 kA of oscillation period 1·2 microsecond. The light for a 4 kA discharge of the shorter period would be less. It appears therefore that the peak light output for a given current is reduced when the time to reach maximum current is sufficiently short.

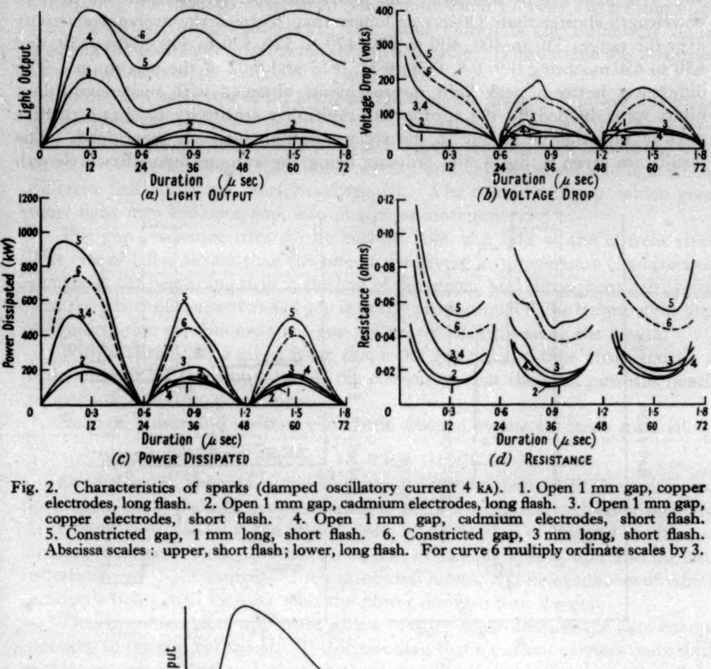


Fig. 2. Characteristics of sparks (damped oscillatory current 4 kA). 1. Open 1 mm gap, copper electrodes, long flash. 2. Open 1 mm gap, cadmium electrodes, long flash. 3. Open 1 mm gap, copper electrodes, short flash. 4. Open 1 mm gap, cadmium electrodes, short flash. 5. Constricted gap, 1 mm long, short flash. 6. Constricted gap, 3 mm long, short flash. Abscissa scales: upper, short flash; lower, long flash. For curve 6 multiply ordinate scales by 3.

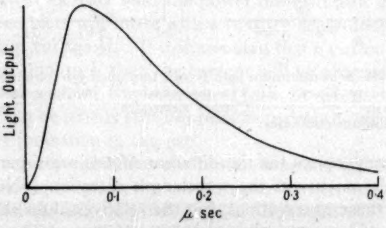


Fig. 3. Light output from open 1 mm gap with damped oscillatory current of 8 kA, oscillation period 0·23 sec (third current peak 4 kA).

The effect of reducing the current by adding damping resistance is greater on the second and subsequent light peaks than on the first. An addition of one tenth of the critical damping resistance reduces the first and second light peaks by about 20% and 50% respectively.

.B
5
2

(iii) *Spectral Distribution of Light*

Some information on the colour of the light emitted from open gaps was obtained by interposing filters between spark and photomultiplier; Ilford colour filters 805, 108, 110, 202, and 205 were used. The transmission of these filters is uniform for long wavelengths but changes from 5% to 50% over a range of about 20 mμ, so that light of short wavelength can be cut off at approximately 400, 470, 520, 570 or 630 mμ. The photomultiplier is insensitive to light of wavelength shorter than 330 mμ or longer than 650 mμ, the average sensitivity over the ranges 330 to 400, 400 to 470, 470 to 520, 520 to 570, 570 to 630, and 630 to 650 mμ being 0·9, 0·8, 0·56, 0·34, 0·13 and 0·02 of the maximum. The differences between peak light measurements obtained with successive colour filters were divided by the appropriate multiplier sensitivity to obtain relative rates of radiation of energy in the six wavelength intervals mentioned. The results are given in fig. 4, the ordinate being the average for each wavelength interval. The characteristics for the different metals are separated in fig. 4 by slight relative displacements in the wavelength direction. No great precision can be expected in these observations, but the relative values shown are unlikely to be out by a factor of 2 except in the range 630 to 650 mμ. The radiation appears to be distributed over the visible range, with a larger proportion of the shorter wavelengths for the short flashes than for the long.

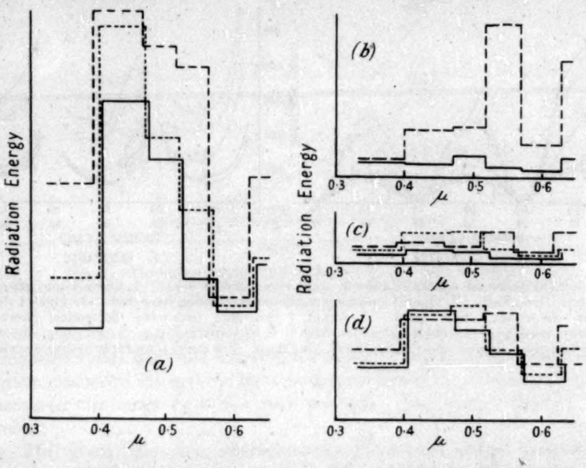


Fig. 4. Spectral distribution of maximum light from 1 mm open gap (obtained by photomultiplier oscillograms and colour filters). (*a*) 4 kA short flash, (*b*) 4 kA long flash, (*c*) 2 kA long flash, (*d*) 2 kA short flash. ———— Copper electrodes - - - - - - - Magnesium electrodes, — — — — — Cadmium electrodes.

Some indication of the relative rates of decay of the different wavelengths was obtained by examining the second, third and fourth light peaks obtained with the different filters. The measurements are not sufficiently precise to show quantitative differences between adjacent wave bands. Over the whole spectrum however it appeared that, with copper and cadmium, light of long wavelength decayed more slowly than that of shorter wavelength, while with magnesium the rate of decay was more uniform.

§ 5. ELECTRICAL CHARACTERISTICS OF SPARK

In fig. 2 (*b*), (*c*) and (*d*) are given electrical characteristics corresponding to the light output curves in fig. 2 (*a*). The voltage given in fig. 2 (*b*) was obtained by subtracting the inductive volt drop (maximum value 17 volts for a 1 mm open gap) from the voltage observed. The value shown is the numerical value, ignoring the polarity reversal in the second half cycle. The current was a damped sinusoid of successive maxima 4, 3·3 and 2·8 kA.

In the first half cycle of the short period flash the maximum power dissipated in the open gap (curves 3 and 4) is about twice that dissipated in the second half cycle—a proportional variation which agrees roughly with that of the light output. With the long period flash (curves 1 and 2) both power and light in the different half cycles are more nearly equal. The constricted gap, which gives more light than the open gap, also dissipates more power.

The gap resistance rises as the current falls, and falls as the current rises. The rate of fall is slower than the rate of rise giving a voltage–time characteristic steeper at the beginning than at the end of the second and subsequent half cycles. The resistance of the constricted gap is nearly proportional to the length of the gap, suggesting that conditions in the gap are nearly uniform along the length.

With a current of a given wave shape the gap voltage–time characteristic is little affected by the amplitude of the current so that the peak power is nearly proportional to the peak current.

Voltage, power and resistance are little affected by the electrode material.

§ 6. CONDITIONS IN THE DISCHARGE GAP

Conditions in the gap, involving many different degrees of excitation and ionization, are likely to be very complicated and not amenable to detailed analysis. It is, however, to be expected that the rate of production of charged particles or 'current carriers' (such as electrons or ions of various kinds), and also the rate of creation of 'light sources' (such as excited atoms, the de-excitation of which produces light), will increase with the power dissipated in the gap.

The current carriers will move with a velocity dependent on the field corresponding to the gap voltage V. If it is assumed that n current carriers move with a velocity proportional to V then the current will be proportional to nV and the ratio of voltage to current, or resistance, will be proportional to $1/n$. The assumption may not be strictly true but resistance data should give some indication of the amount of ionization in the gap.

With the 48 microsecond flash the power–time and light–time characteristics are roughly similar in shape to the current, passing through zero simultaneously. This suggests that the rate of production of 'light sources' and the rate of disappearance through the evolution of light are approximately equal.

With the 1·2 microsecond flash the power dissipated at the same maximum current is more than for the longer flash and the light peak is correspondingly greater. In this case an appreciable time lag occurs between energy dissipation and complete evolution of the resulting light: while the power falls to zero the light from the open gap only falls by two thirds, and the light from the constricted gap scarcely drops at all.

With much shorter flashes the energy may, on account of this time lag, be dissipated before any appreciable amount of light is emitted. The maximum brightness will then depend not on the maximum power but on the total energy dissipated in the gap. For a given current amplitude the power dissipated increases as the duration decreases but the energy, and hence the number of light sources, decreases.* The maximum light, dependent on the maximum number of light sources present at one time, is therefore reduced. There will thus be a limit to the extent to which the duration of a light flash can be reduced, or the maximum brightness for a given current increased, by decreasing the duration of the discharge current.

The greater light from the constricted gap corresponds to a higher power dissipation—the higher resistance is attributed to the maintenance of original gas density be the prevention of expansion.

§ 7. REPEATED SPARKS

A rotary gap having 20 moving contacts mounted on a photographic drum has been used in series with the open flash gap to discharge a condenser of 0·1 μF charged to 8 kv at a repetition frequency of 1000/sec in bursts of 1 second. To minimize the circuit inductance, connection to the rotary spark gap was made by a low inductance concentric cable. A triggered gap, with air blown through the annular space between trigger and anode, has been used in place of the rotary gap, but at 4 kA the heating of the anode is very severe and frequent replacement is necessary.

§ 8. REQUIREMENTS FOR SHORT DURATION SPARKS

The discharge of a capacitance C charged to voltage V through a circuit of inductance L and negligible resistance gives an oscillatory current of maximum value $V(C/L)^{1/2}$ and oscillation period $2\pi\ (LC)^{1/2}$.

For short flashes the product LC must be sufficiently small to give the required duration and $V(C/L)^{1/2}$ must be large enough to give sufficient intensity. The 'duration' cannot be clearly defined but for the most rapid observations the time above half the maximum value is probably a useful guide: for discharges of order of microseconds this is about half a cycle (0·4 μsec for curves 3 and 4 in fig. 2 (*a*) which refers to an oscillation period of 1·2 μsec, and 0·15 μsec in fig. 3 which corresponds to an oscillation period of 0·23 μsec). For very short flashes a pulse of current with a sharp cut off obtained from a transmission line might be considered, but this would require discharging through a resistance equal to $(L/C)^{1/2}$ so that the maximum current obtained would be halved and the peak light would thus be reduced. The light lags behind the current and would not be cut off abruptly. The discharge of a concentrated LC circuit would, therefore, probably give a brighter light source of a given duration than a transmission line circuit of the same L and C. For very short flashes the electrodes should be metals of high boiling point.

For a given duration, i.e. a given product LC, the current and light are greater the lower the inductance L. A minimum inductance is obtained by a circuit consisting virtually of two concentric cylinders as in fig. 5. The condenser is of the tubular form, with paper in oil as dielectric. One electrode is mounted inside the end of a metal cylinder, which pushes over the metal container which is connected to the other foils. The light emerges through a hole in the side of the cylinder.

---

* The statement is supported by the experimental evidence for the 'long' and 'short' flashes given in fig. 2. For very short times a qualitative indication is obtained by assuming that at time T the number n of current carriers and the number N of light sources are proportional to the energy dissipated, E. For a current I the power $P \propto I^2/n \propto dE/dT \propto dn/dT$, whence $n^2 \propto I^2 T$, whence $P \propto I/\sqrt{T}$, and n, N, and $E \propto I\sqrt{T}$.

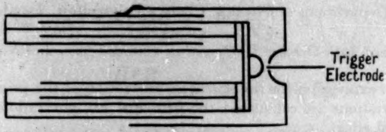


Fig. 5. Low inductance circuit for short time flashes.

An approximation to the concentric cylinder arrangement, using a ring of several standard condensers, has been used by Kovasznay (1949), giving a light source the intensity of which falls to one fiftieth of its maximum in 0·8 microsecond.

§ 9. CONCLUSIONS

A spark carrying a current of 4 kA across an open air gap of 1 mm during the lightly damped oscillatory discharge of a 0·1 μF condenser can provide a point light source with a duration of the order of microseconds. Such a spark can be operated 1000 times in one second by a series rotary spark gap mounted on a camera drum.

Low boiling point electrodes give a light output with a higher maximum and a longer duration, but do not affect the energy dissipated in the gap.

A constricted gap gives light with a higher maximum and longer duration, and dissipates more energy.

The light–time characteristic is a function of the current–time characteristic.

For discharge oscillation periods of one microsecond or more the maximum light intensity for a given maximum current increases as the time to maximum current decreases. For shorter periods, the reverse appears to be the case.

ACKNOWLEDGMENTS

The work described above was carried out in the Electricity Division of the National Physical Laboratory, mainly as part of the research programme of the Laboratory, but partly for the Aeronautical Research Council. It is published by permission of the Director of the Laboratory and with the approval of the Aeronautical Research Council.

The authors desire to acknowledge the assistance rendered by Mr. D. R. Pinkney during part of the investigation.

REFERENCES

Preston, J. S., 1943, *Trans. Illum. Engng. Soc., Lond.*, **8**, No. 7, 132.
Kovasznay, L. S. G., 1949, *Rev. Sci. Instrum.*, **20**, 696.

A Note on the Calculation of Principal Ray Aberration

By R. S. LONGHURST
Department of Physics, Chelsea Polytechnic, London

MS. received 26th October 1950, *and in amended form 2nd November* 1951

ABSTRACT. An expression for the spherical aberration of the principal rays is derived. If the primary aberrations are calculated the principal ray aberration may be found with very little additional effort.

---

§ 1. INTRODUCTION

In discussing the imagery of an optical system it is frequently desirable to calculate the spherical aberration of the principal rays.

If the distortion is corrected for one pair of conjugates and there is no spherical aberration of the principal rays, then it follows that the distortion is corrected for other conjugates. The variation of astigmatism with object distance is also closely connected with the pupil aberration (Merté 1920). In the case of visual instruments, vignetting of the image-forming pencils by the observer's eye pupil may give rise to shadows in the field of view. These shadows are usually reduced to a minimum if there is no spherical aberration of the pupil.

There is an additional effect associated with spherical aberration of the principal rays. If this aberration is large at the first of two widely separated surfaces, the principal ray intersection heights at the second and succeeding surfaces differ considerably from their paraxial values. Consequently the aberrations of these surfaces may differ widely from the first order approximation. This error is in addition to the intrinsic high order aberration of the surfaces concerned and, in certain cases, may be of greater significance. This has been found to be so in the case of Kellner type eyepieces.

§ 2. PRIMARY SPHERICAL ABERRATION OF THE PRINCIPAL RAYS

In most cases it is sufficient to calculate the primary spherical aberration of the principal rays and this calculation can be made to fit into the general primary aberration computing scheme suggested by Hopkins (1950).

In Hopkins' notation, the axial spherical aberration is specified by the term $S_{\mathrm{I}} = A^2 y \Delta(u/N)$ where $A = Ni = N'i'$ and the other symbols have their usual meanings. The corresponding term specifying the spherical aberration of the principal ray is obviously

$$S_{\mathrm{VI}} = B^2 y_{\mathrm{pr}} \Delta(u_{\mathrm{pr}}/N). \qquad \ldots\ldots(1)$$

This expression may be transformed so as to contain only known paraxial quantities.

One has
$$\frac{u_{\mathrm{pr}}}{N} = \frac{\alpha_{\mathrm{pr}}}{N} - \frac{i_{\mathrm{pr}}}{N} = \frac{y_{\mathrm{pr}}}{rN} - \frac{B}{N^2}.$$

Thus,
$$\Delta\left(\frac{u_{\mathrm{pr}}}{N}\right) = \frac{{}_{\mathrm{pr}}}{r}\,\Delta\left(\frac{1}{N}\right) - B\Delta\left(\frac{1}{N^2}\right). \qquad \ldots\ldots(2)$$

Hopkins defines
$$E = \frac{1}{H}\,\frac{y_{\mathrm{pr}}}{y}, \qquad \ldots\ldots(3)$$

where $H = Nu\eta = N'u'\eta'$, the Helmholtz invariant. From (1), (2) and (3) one has

$$S_{\mathrm{VI}} = B^2 HEy\left[\frac{HEy}{r}\Delta\left(\frac{1}{N}\right) - B\Delta\left(\frac{1}{N^2}\right)\right], \qquad \ldots\ldots(4)$$

which is a useful expression if S_{VI} alone is required. If one now writes

$$A\Delta\left(\frac{1}{N^2}\right) = \Delta\left(\frac{i}{N}\right) = \frac{y}{r}\Delta\left(\frac{1}{N}\right) - \Delta\left(\frac{u}{N}\right),$$

eqn. (4) reduces to

$$S_{\mathrm{VI}} = \frac{HEB}{A}[S_{\mathrm{III}} + S_{\mathrm{IV}}(1 + AyE)], \qquad \ldots\ldots(5)$$

where S_{III} and S_{IV} are the wave-front aberration coefficients for astigmatism and field curvature respectively. With the completion of the primary aberration calculations using Hopkins' method, the quantities H, E, B/A, S_{III}, S_{IV} and $(1 + AyE)$ are already available and S_{VI} can be found with very little additional labour.

If the term S_{VI} is large at any surface the value of E at the next surface is altered by virtue of the change in y_{pr}. This will be considerable if the separation between surfaces is large. In addition the distortion at the first surface will alter the value of H. The change δE_2 due to changes δy_{pr_2}, δH_2 is easily seen to be

$$\delta E_2 = \frac{1}{y_2}\left[\frac{H\delta y_{\mathrm{pr}_2} - HE_2 y_2 \delta H_2}{H(H + \delta H_2)}\right]. \qquad \ldots\ldots(6)$$

The angular aberration of the principal ray at the first surface is given by

$$\delta u_{\mathrm{pr}_1}' = \frac{1}{2N_1'}\frac{1}{l_{\mathrm{pr}_1}' u_{\mathrm{pr}_1}'} S_{\mathrm{VI}_1},$$

whence,

$$\delta y_{\mathrm{pr}_2} = -d_1'\delta u_{\mathrm{pr}_1}' = -\frac{d_1'}{2N_1'}\frac{1}{HE_1 y_1} S_{\mathrm{VI}_1}$$

$$= -\frac{d_1'}{2N_1' y_1}\frac{B_1}{A_1}[S_{\mathrm{III}_1} + S_{\mathrm{IV}_1}(1 + A_1 y_1 E_1)].$$

The quantity $d_1'/2N_1'y_1$ is also available from the normal primary aberration calculation.

The distortion is given by $\delta\eta_1' = S_{\mathrm{V}_1}/2N_1'u_1'$, and since $\delta H_2 = N_1'u_1'\delta\eta_1'$, the correction to H is simply $\delta H_2 = \frac{1}{2}S_{\mathrm{V}_1}$. Using (6), these values of δy_{pr_2} and δH_2 give δE_2 and the modified aberrations may be calculated.

It has been found useful to make these calculations when investigating Kellner-type eyepieces where there is a large separation following the steeply curved surface of the field lens. The field curvature of the latter will also affect the aberrations of the eye lens, but this effect is found to be quite small in systems of this type.

ACKNOWLEDGMENTS

This work was carried out under the supervision of Professor L. C. Martin with the aid of a grant from the British Scientific Instrument Research Association.

REFERENCES

Merté, W., 1920, *Z. Phys.*, **1**, 174.
Hopkins, H. H., 1950, *The Wave Theory of Aberrations* (Oxford : Clarendon Press).

Experiments on the Mechanics of Rubber
I: Eversion of a Tube

By A. N. GENT and R. S. RIVLIN
British Rubber Producers' Research Association, Welwyn Garden City, Herts.

Communicated by W. P. Fletcher; *MS. received 12th September* 1951

ABSTRACT. Measurements are made of the changes which take place in the dimensions of rubber tubes when they are turned inside out. These are compared with the values predicted from the theory of large elastic deformations of incompressible isotropic materials.

§ 1. INTRODUCTION

The elastic properties of an ideal, incompressible, highly elastic material which is isotropic in its undeformed state can be specified by a stored-energy function W which must be a function of the strain invariants I_1 and I_2, defined in terms of the principal extension ratios λ_1, λ_2 and λ_3 by $I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$ and $I_2 = \lambda_1^{-2} + \lambda_2^{-2} + \lambda_3^{-2}$. The incompressibility of the material leads to the condition that $\lambda_1\lambda_2\lambda_3 = 1$. For sufficiently small values of the deformation, i.e. of $I_1 - 3$ and $I_2 - 3$, W can be approximated (Rivlin 1949, Rivlin and Saunders 1951) by the expression

$$W = C_1(I_1 - 3) + C_2(I_2 - 3), \qquad \ldots\ldots(1)$$

where C_1 and C_2 are physical constants for the material. This expression for W was first suggested by Mooney (1940) as being applicable as a description of the elastic properties of vulcanized rubber even for large deformations. The experiments of Rivlin and Saunders (1951) indicate, however, that for the particular vulcanizate used, the stored-energy function W is such that $\partial W/\partial I_1$ is constant while $\partial W/\partial I_2$ decreases as I_2 increases from a value of about $\frac{1}{4}\partial W/\partial I_1$ at low values of I_2.

If the assumption is made that the elastic material has a stored-energy function of the Mooney form (1), then the change in dimensions of a circular cylindrical tube when it is everted (i.e. turned inside out) can be calculated (Rivlin 1949). It is found that the changes in the volumes contained by the inner and outer curved surfaces are independent of the value of C_2/C_1 while the change in length is related to C_2/C_1.

If the stored-energy function for the material is not of the Mooney form then, provided $\partial W/\partial I_1$ and $\partial W/\partial I_2$ vary to only a small extent over the range of values of I_1 and I_2 obtaining in the everted tube, we should expect the formulae for the volume changes based on the assumption of a Mooney form for the stored-energy function to be approximately applicable, while the value of C_2/C_1, determined from the change of length on eversion would represent a mean value of $(\partial W/\partial I_2)/(\partial W/\partial I_1)$ over the range of values of I_1 and I_2 obtaining.

In this paper experiments are described in which circular cylindrical tubes of a natural rubber vulcanizate were turned inside out and measurements were made of the changes in the dimensions of the tubes. From these the ratios of the volumes contained by each of the curved surfaces in their deformed and undeformed states are calculated and the results are compared with those calculated on the assumption that the material of the tubes has a stored-energy function of the Mooney form.

Also, from the change in length which occurs on eversion an effective value for C_2/C_1 for the material of the tubes is obtained.

In the experiments, in order to avoid end effects, the measurements were made on the central regions of the tubes.

§ 2. THEORETICAL CONSIDERATIONS

We consider the eversion of a tube of incompressible highly elastic material, which is isotropic in its undeformed state and for which the Mooney form of stored-energy function (1) applies. The external and internal radii of the tube are a_1 and a_2 respectively before deformation. After eversion these radii become $\mu_2 a_2$ and $\mu_1 a_1$ respectively, so that $\mu_2 a_2 > \mu_1 a_1$ and the tube undergoes a change in length. Let λ, which will of course be negative, be the extension ratio. It has been shown by Rivlin (1949) that μ_1, μ_2 and λ are given by the equations

$$\log \frac{\epsilon_1}{\epsilon_2} = \frac{1}{\epsilon_2} - \frac{1}{\epsilon_1}, \qquad \ldots\ldots(2)$$

$$\chi(\epsilon_1 + 1) = \epsilon_2 + 1, \qquad \ldots\ldots(3)$$

$$\frac{2(\lambda^3 - 1)(\lambda C_1 + C_2)(\chi - 1)}{\lambda(C_1 - \lambda^2 C_2)} = (1 + \epsilon_2)\left(\frac{1}{\epsilon_2} - \frac{1}{\epsilon_1}\right), \qquad \ldots\ldots(4)$$

in which the notation $\epsilon_1 = -\lambda\mu_1^2$, $\epsilon_2 = -\lambda\mu_2^2$ and $\chi = a_1^2/a_2^2$ is employed.

Equations (2) and (3) provide a pair of simultaneous equations for the determination of ϵ_1 and ϵ_2. These quantities are therefore independent of C_2/C_1 and depend only on χ. Equation (4) enables us to determine λ in terms of χ and C_2/C_1. Since in the experimental cases with which we shall be concerned in the present paper the changes of length undergone by the tubes on eversion are small, we may write with a sufficient degree of approximation $\lambda = -1 + \alpha$, where α is small compared with unity. A positive value of α corresponds to a decrease in length of the tube. Equation (4) then yields, neglecting terms of the second and higher degrees in α,

$$\frac{C_2}{C_1} = \frac{\alpha(A+5) - (A+2)}{3\alpha(A+1) - (A+2)}, \qquad \ldots\ldots(5)$$

where

$$A = -(1 + \epsilon_1)(1 + \epsilon_2)/2\epsilon_1\epsilon_2. \qquad \ldots\ldots(6)$$

§ 3. EXPERIMENTAL

The experiments were carried out employing three tubes having different values for the ratio a_1/a_2. The tubes which were 6 in. long and of external diameter 1 in. were prepared by a moulding process from a sulphur-less mix having the following composition in parts by weight: natural rubber (pale crepe) 100, *tert*-butyl perbenzoate 2·5. Vulcanization was effected by heating at 140° c for 25 min.

A number of circumferential lines were moulded on the outer surface of each tube by engraving the mould in which they were formed. Measurements were made of the internal and external diameters over the central portions of each tube, both before and after eversion. The change in length of the central portion of each tube was found by measuring the distances between two of the circumferential lines on the tube before and after eversion, by means of a travelling microscope. The measurement was possible in the latter case—in spite of the fact that the circumferential lines were then on the inner surface of the tube—since the sulphur-less vulcanizate employed was sufficiently transparent for the lines to be visible through the tube. It was indeed for this reason that such a vulcanizate was employed.

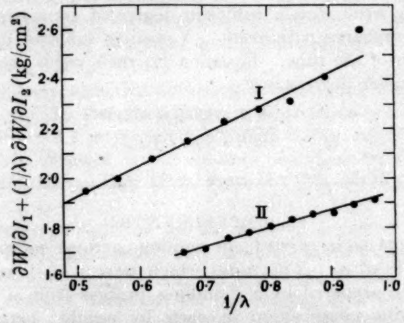
The results of the experiments are given in the table, the notation being that employed in the previous section. We see that there is good agreement between the values of ϵ_1 and ϵ_2 obtained from the measurements and those calculated from eqns. (2) and (3).

Table

a_1 (cm)	a_2 (cm)	$\mu_1 a_1$ (cm)	$\mu_2 a_2$ (cm)	$-\lambda$	ϵ_1 measured	ϵ_2 measured	ϵ_1 calculated	ϵ_2 calculated	$\frac{C_2}{C_1}$
1·27	0·78	0·86	1·30	1·0275	0·47	2·84	0·4583	2·866	0·29
1·27	0·85	0·91	1·31	1·0175	0·52	2·42	0·5110	2·373	0·32
1·27	0·945	$0{\cdot}99_3$	1·30	1·008	0·62	1·91	0·5926	1·876	0·42

From the calculated values of ϵ_1 and ϵ_2 and the measured value of λ, C_2/C_1 is calculated for each tube, by means of eqn. (4), on the assumption that the vulcanizate employed has a stored-energy function of the Mooney form.

The extension of the tube on eversion decreases with the thickness of the tube and therefore the percentage error in the measurement of λ–1 increases. The accuracy of the value obtained for C_2/C_1 is therefore least for the thinnest-walled tube, for which the result is given in the third row of the table. Bearing this in mind it is seen that the results obtained for C_2/C_1 are in approximate agreement with the value obtained for $(\partial W/\partial I_2)/(\partial W/\partial I_1)$ for a different vulcanizate by Rivlin and Saunders (1951).



$\partial W/\partial I_1 + (1/\lambda)\partial W/\partial I_2$ plotted against $1/\lambda$ for simple extension.
I. From experiments of Rivlin and Saunders.
II. For tube used in eversion experiments.

That such an agreement is to be expected, in spite of the very different recipes employed for the vulcanizate in the present experiment and those of Rivlin and Saunders (1951), can be seen from the following considerations.

The measurement of the relation between load and extension ratio for the simple extension of a test piece of vulcanized rubber enables us to obtain the dependence of $\partial W/\partial I_1+(1/\lambda)\partial W/\partial I_2$ on λ for the vulcanizate employed. If this quantity is plotted against $1/\lambda$, a substantially linear relation is obtained for values of λ between 1 and 2 (say). The relation obtained by Rivlin and Saunders for the sulphur-vulcanized compound employed in their experiments is compared in the figure with a similar relation obtained from measurements of the load-extension ratio characteristic of one of the tubes employed in the eversion experiments. Both the slope of the straight line obtained and the intercept on the ordinate are somewhat smaller for the compound used in the present experiment than they are for that employed in the experiments of Rivlin and Saunders. We are thus led to expect that the stored-energy functions for the two vulcanizates will be similar in form and will differ appreciably at any value of I_1 and I_2 only by a multiplying constant.

ACKNOWLEDGMENT

This work forms part of a programme of fundamental research undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

Mooney, M., 1940, *J. Appl. Phys.*, **11,** 582.
Rivlin, R. S., 1949, *Phil. Trans. Roy. Soc.* A, **242,** 173.
Rivlin, R. S., and Saunders, D. W., 1951, *Phil. Trans. Roy. Soc.* A, **243,** 251.

An Interpretation of the Magnetic Properties of some Iron-Oxide Powders

By W. P. OSMOND
Magnetic Tapes Division, Minnesota Mining and Manufacturing Co. Ltd., Slough, Bucks.

MS. received 31*st August* 1951, *and in final form* 16*th October* 1951

ABSTRACT. The observed magnetic properties of fine dispersed powders of the ferromagnetic iron oxides are discussed in the light of modern theories of the magnetization of ferrites and of ferromagnetics containing non-magnetic inclusions. The individual powder particles are shown to be probably of single-domain size, and the relative coercivities of γ-ferric oxide powders of nearly spherical and of markedly acicular grains are qualitatively in agreement with theory, on the grounds of shape anisotropy. After allowance is made for volume concentration and small departures from stoichiometric composition, the ratio of the coercivities of apparently identical acicular powders of γ-ferric oxide and of magnetite is found to be approximately that of the saturation magnetization per ferric ion rather than that of the value per unit volume, and both coercivities are lower than theory would predict. It is concluded that magneto-crystalline anisotropy is also involved. In contrast, relative and actual values of intrinsic induction at the beginning of the approach to saturation are in satisfactory agreement with theory, as are the relative values at remanence although the actual values here are rather low. It is hoped to extend the study to larger departures from stoichiometric composition involving partial transformation to the antiferromagnetic iron oxides α-Fe_2O_3 and FeO.

§ 1. INTRODUCTION

The iron oxides discussed in this paper are those which, in the form of fine powders dispersed in a non-magnetic binder, are used as the magnetic coating of sound-recording tapes. That most commonly employed is the cubic sesquioxide of iron, γ-Fe_2O_3, brown in colour, often referred to as γ-ferric oxide. This is produced by careful oxidation of magnetite Fe_3O_4 which is black. The latter is also used for magnetic tape coatings, but its considerably higher coercivity makes erasing more difficult, and this somewhat offsets its theoretically slightly higher sensitivity. The observed magnetic properties of both oxides will be discussed here for comparison, and an attempt will be made to interpret their differences.

Throughout this paper, the term 'coercivity' will be used to denote the intrinsic coercive force (i.e. the strength of the opposed magnetic field required to reduce the intensity of magnetization to zero) when the previously applied magnetic field has been sufficient virtually to 'saturate' the material, in the sense that no further increase in field produces any measurable increase in residual intensity. Experience shows that the least strength of such an applied field is usually some three times the coercivity as thus defined. It is the latter property which will receive most attention here, since the relative values found for the two forms of oxide appear at first to be anomalous, in contrast to the relative saturation intensities of magnetization. All values given for coercivity and magnetizing field are measured in oersteds and those for magnetic intensity or intrinsic induction in gauss.

§ 2. METHODS OF PRODUCTION AND THEIR EFFECT UPON COERCIVITY

It is beyond the scope of this paper to go into details of the chemical processes of the production of these oxides; only the outlines will be given, with a note of the resulting properties. In all cases γ-ferric oxide is produced by oxidation of magnetite, with little change of shape or size of the powder particles, so that it is the method of producing the parent magnetite which has most apparent effect upon the magnetic properties observed.

The German method for the Magnetophone, fully described in a FIAT report (1947), was by precipitation from a solution of ferrous sulphate and potassium nitrate with an excess of ammonia. The resulting magnetite, when dried and oxidized, produced a red-brown γ-Fe_2O_3 (possibly, from its colour, containing some non-ferromagnetic α-Fe_2O_3). The powder particles seem to have been roughly spherical and of dimensions of the order of $1\,\mu$ or rather less. The measured coercivity of German tapes has been found to be around 90 oersteds. Other tapes coated with precipitated powders of similar particle shape and size have coercivities never exceeding 150 oersteds.

Another method is by reduction of a pigment oxide. The yellow pigment oxides frequently used consist essentially of a monohydrate of Fe_2O_3 with the formula FeO . OH. On heating, the combined hydrogen is driven off, leaving a red powder which in essentials is α-Fe_2O_3. Red pigment oxides are more nearly of this composition to begin with. Reduction of this powder in a suitable atmosphere at a controlled temperature leads to black Fe_3O_4. Re-oxidation at a lower, carefully controlled, temperature gives brown γ-Fe_2O_3. Throughout the cycle from raw material to ferric oxide the powder particles remain markedly acicular. A series of electron-microphotographs shows that there is no noticeable change in particle shape or size between any of the four stages: pigment, dehydrated oxide, magnetite, γ-ferric oxide. The length to width ratio of the individual particles varies between 6 and 10, with individual lengths ranging from $0{\cdot}1\,\mu$ to at most $1{\cdot}5\,\mu$, and there is some evidence that the larger grains, more particularly in the two ferromagnetic oxides, may consist of smaller grains intimately held together. The coercivities of tapes coated with such powders are of the order of 310–330 (magnetite) and 215–230 (γ-ferric oxide). The ratio of the latter to the former is thus approximately 0·7, both being several multiples of the values found for tapes coated with powders derived from precipitated magnetite.

§ 3. BASIC CAUSES OF THE DIFFERENT COERCIVITY VALUES

(i) *General Survey*

Relatively small variations in magnetic properties would be expected from the presence of small variable amounts of impurities in the oxides, as well as from small departures from stoichiometric chemical composition, both of which effects can be kept within narrow limits. The present aim is to study the major measured variations (*a*) between different powder forms of the same oxide and (*b*) between different oxides of the same powder form, and these effects will at present be neglected, as will the effect of mutual interactions between grains, the ‘ apparent density’ effect of Néel (1947 a) and differences in magneto-crystalline anisotropy, since both oxides have virtually the same spinel-cubic lattice (Wells 1945, Verwey and Heilmann 1947, Néel 1948 a).

(ii) *Major Differences between Powders of the same Oxide*

The outstanding example of differences between powders of the same oxide is provided by coercivity figures for German and similar tapes, coated with γ-ferric oxide formed from precipitated magnetite (90–150), compared with those for tapes coated with acicular particles of the same oxide derived from magnetite produced by reduction (215–230). The ratio of these coercivities is of the order of 0·4 to (at most) 0·7. Although the relative volume concentration of oxide in the tape coatings is not known, it is highly probable that the difference in average shape factor of the powder particles is the dominant cause of so large a difference.

Essentially identical formulae for the mean coercivity of a ferromagnetic consisting of an assemblage of single-domain particles of random orientation were given almost simultaneously by Néel (1947 a) and Stoner and Wohlfarth (1947, 1948). From these formulae the coercivity is proportional to the saturation magnetization I_0 and to the difference between the demagnetizing factors along the equatorial and polar axes of the ellipsoid of revolution to which a powder particle can be assumed to approximate. Since this difference increases with increasing length/width ratio, the coercivity should be greater for an assemblage of long, thin particles than for a precisely similar assemblage of more nearly spherical particles of the same saturation intensity per unit volume of ferromagnetic substance.

For the ordinary cubic ferromagnetic metals iron and nickel Stoner and Wohlfarth (1948, p. 635) calculate that the maximum permissible length of a single-domain particle of length/width ratio 10 is $6{\cdot}12 \times 10^{-5}$ cm ($0{\cdot}612\,\mu$) for iron and $25{\cdot}2 \times 10^{-5}$ cm ($2{\cdot}52\,\mu$) for nickel, the permissible diameters for spheres being respectively $1{\cdot}5 \times 10^{-6}$ cm ($0{\cdot}015\,\mu$) and $6{\cdot}24 \times 10^{-6}$ cm ($0{\cdot}0624\,\mu$). Their formula (6.11), admittedly only approximate, from which these values are derived can be written:

$$b^2 \ngtr \frac{\pi}{20}\,\frac{\boldsymbol{k}\theta}{D_{\mathrm{a}}}\left(\frac{a_1{}^3}{q}\right)^{5/3}\frac{1}{q}\,\frac{1}{\mu_{\mathrm{B}}{}^2}, \qquad \ldots\ldots(1)$$

where b is the equatorial radius of the spheroid, D_{a} the demagnetization coefficient along the polar axis ($D_{\mathrm{a}} = N_{\mathrm{a}}/4\pi$, where N_{a} is the more usual demagnetization coefficient), **k** is Boltzmann's constant, μ_B the Bohr magneton, θ the Curie temperature in °K, a_1^3 the volume per atom, and q the number of effective electrons per atom, so that a_1^3/q is the volume per effective electron.

We will apply this formula to these oxides which consist of assemblages of large oxygen atoms and small iron atoms.

For magnetite $\theta = 848°$ K (Néel 1948 a), and the crystal structure is inversed spinel (Verwey and Heilmann 1947), the unit cell being a cube of edge 8·39 Å and containing eight molecules. Per molecule the volume is therefore $(4{\cdot}195 \times 10^{-8})^3$ cm^3, and the average arrangement is one Fe^{3+} ion in a tetrahedral interstice (A site) between the close-packed oxygen ions with another Fe^{3+} and one Fe^{2+} in octahedral interstices (B sites). Néel (1948 a) explains the saturation magnetization of magnetite as being entirely due to this one Fe^{2+} ion per molecule with four effective electrons, the two Fe^{3+} ions being arranged with antiparallel magnetic spins, hence with mutual cancellation of magnetic moments. This conclusion has been supported by Gorter (1950) and other recent writers, but like Weiss and Forrer (1929) they found a somewhat higher value of magnetic intensity than $4\mu_B$ per molecule, probably owing to an incompletely quenched orbital moment, hence a splitting factor g rather greater than 2. However, the variation is insufficient vitally to affect the present calculations.

We thus obtain for the volume per effective electron

$$a_1^3/q = (4{\cdot}195 \times 10^{-8})^3/4, \qquad \ldots\ldots(2)$$

a result which is unaffected by the number of atoms per molecule. There are actually seven of course, four large O and three small Fe, but it might be considered that only the iron atoms (or possibly even only the one Fe^{2+} ion) should be considered. However, since 'superexchange' interactions between the iron atoms via the oxygen ions are probably involved (Néel 1948 a), the presence of the latter cannot be ignored magnetically, even if it were legitimate to ignore their material existence. We write therefore 1·75 for the value of the factor $1/q$ in (1), while noting that the results obtained may possibly be rather too high.

Substituting all these values into (1) we obtain

$$b \ngtr 2{\cdot}2 \times 10^{-6}\, D_a^{-1/2} \qquad \ldots\ldots(3)$$

compared with $4{\cdot}35 \times 10^{-7} D_a^{-1/2}$ for Fe and $1{\cdot}80 \times 10^{-6} D_a^{-1/2}$ for Ni (Stoner and Wohlfarth 1948, eqn. 6.12). It follows that the maximum permissible dimensions for single-domain particles of magnetite are probably of the same order as those for nickel and possibly somewhat larger.

The Curie temperature of pure γ-ferric oxide is 675° C (Michel, Chaudron and Bénard 1951) although rarely reached during heating, owing to transformation to the rhombohedral antiferromagnetic α-form for which the Curie point (transition point), when showing feeble ferromagnetism, is also 675° C (Néel 1949, Chevallier 1951). When the oxide is stabilized by 'foreign' atoms the Curie temperature is lowered. In assuming then that $\theta = 675°\text{C} = 948°\text{K}$ the error may be somewhat high, leading to rather too high a value for b.

From the approximately linear relationship between the iron atomic concentration and lattice constant a of a cube containing 32 O atoms described by Wells (1945) for the cubic oxides of iron, we calculate that the unit cell of γ-ferric oxide is a cube of edge approximately 8·28 Å. This is a 'defect' spinel structure, and the unit cell contains on the average $10\frac{2}{3}$ molecules, the average arrangement per molecule being $\frac{3}{4}$ of an Fe^{3+} ion in a tetrahedral hole, balanced by $\frac{3}{4}$ of an Fe^{3+} ion in an octahedral interstice, with $\frac{1}{2}$ of an Fe^{3+} ion (providing the net effective 2·5 electrons per molecule) also in an octahedral site (Verwey and Heilmann 1947, Néel 1948 a). Hence the volume per molecule is $0{\cdot}75\,(4{\cdot}14\times10^{-8})^3$ cm³ while for the volume per effective electron

$$a_1{}^3/q=0{\cdot}3\,(4{\cdot}14\times10^{-8})^3 \qquad \ldots\ldots(4)$$

and $1/q=5/2{\cdot}5=2$, since there are five atoms in the molecule.

From these figures we obtain

$$b \ngtr 2{\cdot}803\times10^{-6}\,D_a{}^{-1/2} \qquad \ldots\ldots(5)$$

which is somewhat larger than the value obtained for magnetite, and hence still larger than that for nickel. Furthermore, had the (smaller) value of θ used for magnetite been taken here we should still have obtained a value for the upper limit of b as high as $2{\cdot}653\times10^{-6}\,D_a{}^{-1/2}$.

Since, as stated earlier, the dimensions of the largest powder particles as revealed by electron microphotographs are less than the upper limits for single-domain particles of nickel, there seems every reason to assume that these oxide powder particles are really single domains.

From table 7 and fig. 8 of Stoner and Wohlfarth (1948, p. 625) it can be seen that (assuming identical circumstances otherwise) a reduction of the coercivity to 0·4 times its original value should arise from reducing the mean dimension ratio from between 6 and 10 to between 1·6 and 1·7, the corresponding figures for reduction to 0·7 being between 2·5 and 3·0. It is clear, therefore, that the single-domain particle theory can account, qualitatively at least, for the relative measured coercivities of these tapes coated with different powder forms of essentially the same oxide, in terms of shape anisotropy alone.

(iii) *Major Differences between Different Oxides of the same Powder Form*

The principal example of a difference between different oxides of the same powder form is provided by the values found for apparently exactly similar powders of magnetite and of the γ-ferric oxide produced from it. Given identical powders and powder dispersions the magnetic properties would be expected to vary in strict accordance with the relative values of saturation magnetization per unit volume. As will be shown later, the relative measured values of residual intrinsic induction, as well as the apparent saturation values, can be satisfactorily explained in this way. However, the relative measured coercivities present a marked anomaly when compared with the results of theory, and this discrepancy will be fully treated now.

Saturation intensity I_0. We first need an estimate of the relative values of I_0, the saturation magnetization per unit volume. For this we use the lattice constants and saturation magnetization values per molecule given in § 3 (i), but noting that these are probably strictly applicable only at very low temperatures. However, the curve given by Néel (1948 a, fig. 18) for the thermal variation of spontaneous magnetization of magnetite shows that, at room temperature, roughly one-third of the interval from 0° K to the Curie point, the value is still above 90% of that near the absolute zero. Thus at 0° K, eight molecules of Fe_3O_4 occupy $(8{\cdot}39)^3\times10^{-24}$ cm³ with total magnetic intensity $32\,\mu_B$. Hence

$$I_0=32\times9{\cdot}25\times10^3/(8{\cdot}39)^3=502. \qquad \ldots\ldots(6)$$

At room temperature, assuming some 95% of this value, we obtain as a fair estimate $I_0 = 480$. A similar value would be obtained by taking $4{\cdot}2\,\mu_B$ (Gorter 1950) for the magnetic moment of the molecule and assuming only 90% of this value at room temperature. For comparison we have $4\pi I_0 = 5400$, $I_0 = 430$, given, as a representative value only, by Stoner (1934), while Birks (1950), Bickford (1950), and Domenicali (1950) give values ranging from 450 to 470.

For γ-ferric oxide at 0° K, $10\frac{2}{3}$ molecules occupy $(8{\cdot}28)^3 \times 10^{-24}$ cm^3 with total magnetic intensity $10\frac{2}{3} \times 2{\cdot}5\,\mu_B$. Hence

$$I_0 = 26{\cdot}67 \times 9{\cdot}25 \times 10^3/(8{\cdot}28)^3 = 435. \qquad \ldots\ldots(7)$$

At room temperature it is reasonable to assume a similar reduction to that for magnetite, viz. to 90–95% of this value (Néel gives no curve of thermal variation of intensity for this oxide, the difficulty being that it becomes unstable around 700° K and transforms to the antiferromagnetic α-form at about 920° K). We should therefore expect a value for I_0 of between 392 and 414, and if we assumed 95% of the figure obtained by using Weiss and Forrer's $2{\cdot}39\,\mu_B$ per molecule (Néel 1948 a) we should have 395·7. A reasonable estimate therefore is $I_0 = 400$, which is five-sixths of the estimated value for magnetite.

It follows that, for identical agglomerations of powders of exactly similar grains of γ-ferric oxide and magnetite, the mean coercivity ${}_IH_{\mathrm{cm}}$ of the former should be some $\frac{5}{6}$ (0·83) times that of the latter. As mentioned in §2, the measured ratio for tapes is usually of the order of 0·7, and the same is true for powders alone.

Mutual interactions. The mutual interactions of the powder can be estimated from the 'apparent density' formula given by Néel (1947 a) which can be written

$${}_IH_{\mathrm{cm}} = (1 - \rho_a/\rho_0)\,({}_IH_{\mathrm{cm}})_0, \qquad \ldots\ldots(8)$$

where ρ_0 is the true density of the ferromagnetic part of the substance, ρ_a the 'apparent' density and $({}_IH_{\mathrm{cm}})_0$ the theoretical mean coercivity for a powder with no interactions (i.e. 'apparent' density $\rho_a \to 0$, representing a state of infinite dilution).

If V is the volume concentration of non-ferromagnetic inclusions, $1 - V$ is that of the ferromagnetic itself; hence $\rho_0(1 - V) = \rho_a$, so that the formula becomes

$${}_IH_{\mathrm{cm}} = V({}_IH_{\mathrm{cm}})_0. \qquad \ldots\ldots(9)$$

This relationship has been verified with very fair accuracy by Weil (1947, 1948) for powders in which shape anisotropy is fairly conclusively the dominating factor influencing the coercivity; it is, therefore, reasonable to assume its general applicability to the fine powders now being considered.

The active coating of a magnetic recording tape consists of a dispersion of the fine powder grains in a non-ferromagnetic binder, the constituents of the latter having a mean density of about 0·9. We need an estimate of the true density of the ferromagnetic oxides.

For magnetite one mole weighs 231·55 g and eight molecules occupy $(8{\cdot}39)^3 \times 10^{-24}$ cm^3; hence $\rho_0 = 5{\cdot}206$ (cf. Kaye and Laby (1944): 5 to 5·4). Similarly for γ-ferric oxide one mole weighs 159·7 g and $10\frac{2}{3}$ molecules occupy $(8{\cdot}28)^3 \times 10^{-24}$ cm^3; hence $\rho_0 = 4{\cdot}984$. Taking 2 as a convenient typical value of the weight ratio of oxide to binder we obtain the following results for magnetite:

Density ratio of oxide to binder $5{\cdot}206/0{\cdot}9 = 5{\cdot}784$
Weight ratio of oxide to binder 2
Volume ratio of oxide to binder 1/2·892
Hence volume concentration of oxide $1 - V = 1/3{\cdot}892 = 0{\cdot}2569$
volume concentration of binder $V = 0{\cdot}7431$.

Substituting in (9) we have

$$ {}_I H_{\mathrm{cm}} = 0{\cdot}7431\ ({}_I H_{\mathrm{cm}})_0 \quad \text{(magnetite).} \qquad \ldots\ldots(10) $$

γ-ferric oxide. By exactly similar reasoning we obtain:

$$ {}_I H_{\mathrm{cm}} = 0{\cdot}7346\ ({}_I H_{\mathrm{cm}})_0 \quad (\gamma\text{-ferric oxide).} \qquad \ldots\ldots(11) $$

The ratio of the measured coercivity of the latter oxide to that of the former should therefore be 98·86% of the ratio for infinite dilution rather than 0·7 as frequently found in practice.

It is difficult to estimate the volume concentration of air in powders examined alone. However well they are pressed down in small test tubes inserted in the magnetizing coil of the test gear, the agitation of the rapidly alternating field causes some loosening of the packing, and measured coercivities differ little from those found for tapes. In general there appears to be a slight reduction, up to say 3%, representing an increase in powder concentration of 8 to 9%.

Chemical composition. Finally we can study the probable effect on I_0, and hence on coercivity, of the small departures from stoichiometric composition usually found in practice. The magnetite powder produced by the reduction process is stabilized after cooling by allowing a slight surface oxidation in air. Chemical analysis then gives an estimated FeO content of 29 to 30% by weight instead of 31·03% as for stoichiometric Fe_3O_4. Assuming that the basic spinel structure is maintained, and neglecting the extremely small volume change, we can adapt Néel's theory as follows.

Let the molecular ratio of FeO to Fe_2O_3 be p/q, and let the estimated weight content of FeO be x%. Then

$$ 71{\cdot}85\,p/159{\cdot}7\,q = x/(100-x) \quad \text{or} \quad p/q = 2{\cdot}223\,x/(100-x). \qquad \ldots\ldots(12) $$

With $x = 29{\cdot}5$ (a typical tolerated value) $p/q = 93{\cdot}03\%$, whereas for stoichiometric magnetite the value is 100%. The simplest, although slightly inaccurate, assumption to make is that only 93% of the B sites normally occupied by Fe^{2+} ions are so filled, so that I_0 is no more than 93% of the value for pure magnetite. This method is probably too crude for considerable departures from ideal composition, but for relatively small variations it should give results of the right order of magnitude.

Again, it is found difficult to oxidize magnetite powders right down to γ-Fe_2O_3, a tolerance of 1% FeO by weight being often found. Using the same method (and, with this small percentage, with much less inaccuracy) we find $p/q = 2{\cdot}25\%$. Hence on the average for every 400 molecules of Fe_2O_3 (800 Fe^{3+} ions) there are nine Fe^{2+} ions, and the simplest assumption is that the latter occupy B sites which would be vacant in pure γ-ferric oxide. On this basis there should be an increase in I_0 amounting to $36\,\mu_B$ per 400 molecules, i.e. an increase of 3·6%.

We should thus expect the ratio of the coercivity of a tape coated with the latter oxide to that of a precisely similar tape coated with the former (assuming the same powder form in each case) to be $103{\cdot}6/93 = 1{\cdot}114$ times the value that would be found with stoichiometric oxides. In spite of the somewhat limited accuracy of the calculations, the trend must be in the direction shown, so that the observed ratio of approximately 0·7 would appear to be considerably *higher* than that proper to stoichiometric compositions.

Examples from actual measurements. Two examples from the complete production cycle of actual powders bring out the last point very clearly.

(*a*) Raw material a red pigment oxide.
After reduction : $_IH_c = 315$, FeO content 28%
After oxidation : $_IH_c = 240$, FeO content 1·76%
Ratio of measured coercivities 240/315 = 0·76.

By the method just described we calculate the theoretical coercivities of exactly similar powders of pure oxides to be : magnetite, 364·2 ; γ-ferric oxide, 225·6 ; ratio 0·619.

(*b*) Raw material a yellow pigment oxide.
After reduction : $_IH_c = 315$, FeO content 29·24%
After oxidation : $_IH_c = 230$, FeO content 2%
Ratio of measured coercivities 230/315 = 0·73.

Hence for similar pure oxides the theoretical coercivities are : magnetite, 342·9 ; γ-ferric oxide, 214·4 ; ratio 0·625.

A considerable number of such examples have been studied, all yielding comparable results. The ratio of theoretical coercivities is always well below 0·7, almost invariably between 0·61 and 0·64 and frequently between 0·62 and 0·63.

(*c*) These principles have been applied to two actual tapes, the first, a magnetite coated tape of measured coercivity 325, and the second, one coated with γ-ferric oxide of measured coercivity 225, the coercivity ratio being 0·692. We will assume the same volume concentrations as those used before, and FeO contents 29·5 and 1% respectively (the actual values are not known, but the results obtained with these figures must be of the correct order of magnitude).

For the magnetite tape the probable coercivity of an exactly similar stoichiometric powder without interactions is calculated as $325/(0{\cdot}7431 \times 0{\cdot}93) = 470$. For the other the calculation is $225/(0{\cdot}7346 \times 1{\cdot}036) = 296$. The coercivity ratio is now 0·629.

Summary. There is good evidence that the powder grains of magnetite and of the γ-ferric oxide produced from it have almost identical mean shape factors, yet after taking due account of their relative densities, both absolute and apparent, and of usual small departures from stoichiometric composition, we can by no means attain a ratio of mean intrinsic coercivities as high as that of reasonably estimated spontaneous magnetizations *per unit volume* of the substance of the grains; yet, as will be shown in the next section, the latter are qualitatively well substantiated by experimental observations of intrinsic induction.

In fact, the ratio of the coercivities calculated for exactly similar powders of pure oxides without interactions comes out to very approximately $\frac{5}{8}$, which is the ratio of Néel's values of the saturation intensity *per molecule* of each oxide. This is of course the ratio of the net intensity per molecule of Fe_2O_3 present in the molecule of each oxide, and hence of the net intensity per individual ferric ion.

This is the anomaly referred to earlier, and no adequate explanation can be offered for so surprising a result on the assumption of shape anisotropy alone. In this connection it is of interest to note that the figure of 470 oersteds calculated in the last sub-section for stoichiometric magnetite without interactions corresponds to a mean length/width ratio of about 1·55, which is very different from the value of 6 to 10 estimated from electron-microphotographs, and the value derived from the estimated 296 oersteds for γ-ferric oxide is lower still.

It thus appears that some other important factor besides the possibly more important shape anisotropy ought to be taken into account, and that the effect on γ-ferric oxide is more pronounced than that on magnetite.

Magneto-crystalline anisotropy. Stoner and Wohlfarth (1948, p. 637) state that, for cubic crystals in which magneto-crystalline anisotropy is the predominating factor affecting rotational hysteresis, the coercivity will be given by $\alpha K/I_0$, where α depends upon direction and has a maximum value of 2, and K is the numerical value of the anisotropy coefficient. The only values of K for both these oxides that have been found are those due to Birks (1950), obtained from an expression for the internal (anisotropy) field $H_N = 2K/I_0$, hence precisely of the same form as the above expression for the coercivity. The values for I_0 used by Birks, derived from Weiss and Forrer's (1929) values at room temperature, are respectively 450 and 350, compared with the 480 and 400 assumed in the present calculations which are based on later work, the ratio of the values for γ-ferric oxide and magnetite being much the same in each case (0·78 compared with 0·83). Birks finds $|K| = 1{\cdot}3 \times 10^5$ erg/cm for magnetite (cf. Bickford (1950) $1{\cdot}1 \times 10^5$) and $4{\cdot}7 \times 10^4$ for γ-ferric oxide, the ratio of the latter to the former being 0·356. If we use Birks' values of H_N but the higher values of I_0 the ratio becomes 0·381. However, for comparison of theoretical coercivities for similar orientations of easy axes we need to compare the values of K/I_0, so that we come back finally to comparison of the values of H_N found experimentally by Birks. These are 270 for γ-ferric oxide and 590 for magnetite, with a ratio of 0·458.

It seems clear therefore that the influence of magneto-crystalline anisotropy in both oxides may well account for the reduction in the ratio of their coercivities from the expected $\frac{5}{6}$ for shape anisotropy alone to the observed $\frac{5}{8}$, as well as for the relatively low values of the actual coercivities. Taking $H_N = 590 \simeq 1{\cdot}23\, I_0$ for magnetite, and $H_N = 270 \simeq 0{\cdot}675\, I_0$ for γ-ferric oxide, we can see that equal weighting for the effects of shape anisotropy and magneto-crystalline anisotropy might well result in a greater proportional lowering of the coercivity of γ-ferric oxide, which is in keeping with the findings of the final paragraph of the last sub-section. It would not be profitable, however, to attempt numerical estimates of the relative importance of the two effects from the rather indefinite information at present available.

Strain anisotropy. With so many unknowns it is even more impossible to estimate the effect of any differences in strain anisotropy. Although internal strains may be expected to be of importance, as in nickel, to which these spinel oxides bear close resemblances in their magnetic properties, it seems unlikely that there should be a consistent difference between tape coatings of the two types of oxides. Variation in lattice strains due to the presence in magnetite of the larger Fe^{2+} ions, as well as to the presence of more Fe ions for the same oxygen content in the unit cell, are probably accounted for in differences of magneto-crystalline anisotropy.

§ 4. MAGNETIZATION CURVES FOR COATED TAPES

(i) *Estimated Saturation Intensities*

If I_0 is the saturation magnetization of a ferromagnetic substance itself, the effect of the presence of non-ferromagnetic inclusions of volume concentration V is to reduce the saturation intensity of the compound to a value $I_m = (1 - V)I_0$.

Using the values calculated in the last section, we find as estimates of the order of magnitude to be found in these tapes:

$$\left.\begin{array}{ll}\text{Magnetite tape:} & I_{\mathrm{m}}=0{\cdot}2569\times 480=123{\cdot}3,\\ \gamma\text{-ferric oxide tape:} & I_{\mathrm{m}}=0{\cdot}2654\times 400=106{\cdot}2.\end{array}\right\} \quad \ldots\ldots(13)$$

These figures refer to pure oxides; with the small departures from stoichiometric composition (29·5% and 1% respectively) considered in the last section the estimates become 114·7 and 110. In terms of intrinsic induction $B_{\mathrm{i}}=4\pi I$, we then have for the expected order of magnitude:

$$\left.\begin{array}{ll}\text{Magnetite tape:} & B_{\mathrm{im}}=1441,\\ \gamma\text{-ferric oxide tape:} & B_{\mathrm{im}}=1382.\end{array}\right\} \quad \ldots\ldots(14)$$

(ii) *The Approach to Saturation*

In practice tests are made on these tapes by observing cathode-ray oscilloscope traces with 50 c/s alternating magnetic fields, the method being essentially the same as that described in an earlier paper (Hobson, Chatt and Osmond 1947). The maximum peak field is generally approximately 1000 oersteds, so that true saturation is not reached. For comparison of theoretical and experimental figures we need an estimate of the values of B_{i} for both tapes at $H=1000$.

A law of approach to saturation for a ferromagnetic containing cavities or non-ferromagnetic inclusions has been given by Néel (1948 b). With the same symbolism as that already used in this paper, Stoner (1950, p. 143) gives this law in the form:

$$1-\frac{I}{I_{\mathrm{m}}}=\frac{V}{2(1-V)}\left[\frac{2+3\alpha}{4(1+\alpha)^{1/2}}\ln\left\{\frac{(1+\alpha)^{1/2}+1}{(1+\alpha)^{1/2}-1}\right\}-\frac{3}{2}\right], \quad \ldots\ldots(15)$$

where $\alpha=H/4\pi I_{\mathrm{m}}=H/B_{\mathrm{im}}$.

We will now apply this law to the two tape coatings considered in the last sub-section, using the final values given there as typical of slightly non-stoichiometric oxides with usual volume concentrations, in order to estimate the values of I/I_{m} at $H=1000$.

Magnetite tape: $\alpha=1000/1441=0{\cdot}694$, $V=0{\cdot}7431$. Hence $I/I_{\mathrm{m}}=0{\cdot}865$ and $B_{\mathrm{i}}=1247$.

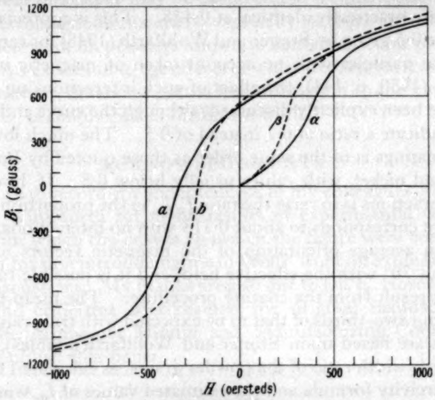
γ-ferric oxide tape: $\alpha=1000/1382=0{\cdot}724$, $V=0{\cdot}7346$. Hence $I/I_{\mathrm{m}}=0{\cdot}877$ and $B_{\mathrm{i}}=1212$.

These values are nearly the same, and a slight increase in the degree of departure from stoichiometric composition of both oxides can easily result in a higher figure for a γ-ferric oxide tape than for one coated with magnetite (e.g. for 29 and 2% FeO we obtain respectively 1224 and 1246 by the same method). Such practically coincident peak values at $H\simeq 1000$ are the usual experience in comparing tapes, together with virtually identical values of residual induction, although it must be borne in mind that small differences in coating thicknesses will have some effect on the observed values, which are really measurements of flux rather than flux densities.

(iii) *Experimental Curves*

With careful compensation for the air space in the search coil containing the specimen, the trace obtained directly on the cathode-ray oscilloscope is one showing the variation with H of the time rate of change of intensity dI/dt. By means of an integrating circuit (time integration) the hysteresis loop I (or $B_i = 4\pi I$) vs H is displayed. The integration can be made very accurate, but use of this circuit causes considerable attenuation of the trace.

To obtain the curves shown in the figure both traces were photographed for each of two tapes, one coated with magnetite and the other with γ-ferric oxide, and careful tracings of projected enlargements were made on squared paper. In order to avoid possible errors in the integrating circuit the differential curves were first converted into true differential susceptibility curves χ_d vs H (where $\chi_d = dI/dH$) by multiplying each ordinate by sec ωt on the assumption of a truly sinusoidal alternating field $H = H_m \sin \omega t$. The integration was then performed by summation of areas $\chi_d dH$, but was only carried out between $H = -0{\cdot}9\,H_m$ and $H = +0{\cdot}9\,H_m$, thus avoiding the indeterminacy resulting from the extremely large values of sec ωt as $\omega t \rightarrow \pi/2$ and from the rounded tips of the trace as the spot on the screen of the oscilloscope changes the direction of its horizontal travel. Since H_m was approximately 1175 the integration extended between $H \simeq \pm 1050$. The resulting sum represents the change in I between $-0{\cdot}9\,H_m$ and $+0{\cdot}9\,H_m$ or vice versa, and in order to fix the point $I = 0$ allowance was made for the small difference (5 to 6%) between the ordinates of the rising and falling curves of the photographed hysteresis loop at each of these extremes. The result is now the upper portion of the hysteresis loop, and on reducing the scale to that of the photographic traces the fit was found to be exact for the magnetite tape and with purely negligible differences for the other. Both methods of integration are therefore considered to be of good accuracy. Finally, these integrated curves were replotted as in the figure, the scales being calibrated from the magnetite tape of known remanence (550) and coercivity (325) kindly supplied by the Minnesota Mining and Manufacturing Company of America.



a, hysteresis loop; α, initial magnetization curve for tape coated with magnetite; *b* and β, similar curves for a coating of γ-ferric oxide.

For this tape B_i at $H = 1000$ as read from curve a is 1059 and 1125 with increasing and decreasing fields respectively; the mean value, $B_i = 1092$ may therefore with tolerable accuracy be taken as the corresponding value on the initial magnetization curve. From curve *b* the corresponding figures for the γ-ferric oxide tape are 1093, 1157 and 1125 with $B_{\mathrm{r}} = 565$ and ${}_IH_{\mathrm{c}} = 227$. The ratio of these mean values is almost identical with that of the theoretical values found before for the greater departures from stoichiometric composition, but the figures are some 10% lower, although they would be the same if the tape thicknesses were 0·62 and 0·63 mil respectively (1 mil $= 10^{-3}$ in.) instead of the 0·7 mil assumed in calculating the scale in gauss in the figure. In view of the empirical nature of the assumptions made this agreement between measured and theoretical values is qualitatively very satisfactory.

The curves labelled α and β in the figure are derived from the means of the B_{I} values for the two branches of the hysteresis loop at each value of H, and may thus be considered as reasonable approximations to the initial magnetization curves for the two tapes. For closer comparison with theory much higher fields are necessary than are at present possible with this particular method of testing. However, both curves continue to fit Néel's formula (15) quite reasonably at values of H somewhat lower than 1000, more particularly that for γ-ferric oxide.

(iv) *Remanence Values*

From both curves of the figure the ratio of remanence I_{r} to the value of I at $H = 1000$ is approximately 0·502. Using the estimated saturation values for B_{Im}, the ratios $I_{\mathrm{r}}/I_{\mathrm{m}}$ are practically identical at 0·438. This is appreciably less than the theoretical value 0·5 given by Stoner and Wohlfarth (1948) for random orientation of single-domain particles with no account taken of magnetic interactions. As Stoner remarks (1950, p. 144), the effect of such interactions on remanence does not seem to have been explicitly discussed, although the curve and figures given by Néel (1947 b) indicate a ratio of 0·7 instead of 0·5. The much lower figure found for these tape coatings is of the same order as those quoted by Bozorth (1948) for annealed iron and nickel, with values usually below 0·5. If, however, the effect of magnetic interactions is to raise the ratio $I_{\mathrm{r}}/I_{\mathrm{m}}$ in the proportion of 7 : 5, then the 0·438 found here corresponds to about 0·313 with no interactions. Such a figure is suitable to an average orientation of the magnetic vectors of the individual particles of about 70° with the effective field, and it is possible that something of this nature may result from the coating procedure. The mean coercivity would then be only about two-thirds of that to be expected with truly random orientation (these estimates are based upon Stoner and Wohlfarth's tables), and as a result the average length/width ratio of the powder grains, as calculated from Stoner and Wohlfarth's coercivity formula and the estimated values of I_0, would be considerably nearer the values of 6 to 10 shown by electron microphotographs than are the values found when using the formula for random orientation. This suggestion is of course purely speculative.

Summarizing, the experimental curves show very similar behaviour with decreasing field from relatively near saturation down to remanence, the ratio of the intensities of the two oxides being of the expected order of $\frac{5}{8}$ when reasonable allowance is made for chemical composition and volume concentration. It is only when the field changes sign that the change in general shape of the curves begins, that for γ-ferric oxide becoming considerably the steeper as the relatively low coercivity point is approached. These curves are quite typical, and, apart from the use of the calibrated magnetite tape, were not specially selected.

§ 5. CONCLUSION

In this paper an attempt has been made to relate the observed magnetic characteristics of dispersions of fine powders of the ferromagnetic oxides of iron to modern theories of the magnetization of materials consisting in the main of single-domain ferromagnetic particles embedded in a non-magnetic matrix. There is strong evidence that the powder particles considered are in fact single ferromagnetic domains.

Measured values of intrinsic induction at the beginning of the approach to saturation have been found to be in good qualitative agreement with theoretical values based on reasonable estimates of concentration and saturation magnetic intensity of the individual oxide particles. The relative remanence values found for the two basic types of oxide are also consistent with the respective estimates, but are lower than theory would predict. The measured coercivities of different powder forms of the same oxide are related in a way which can be satisfactorily accounted for in terms of shape anisotropy alone, but those of similar powders of the two basic oxides show an anomalous relationship from this point of view. It seems clear that magneto-crystalline anisotropy plays an important part and that it causes a larger relative reduction in the coercivity of γ-ferric oxide than in that of magnetite.

No more than qualitative agreement between observation and theory is claimed, but it is hoped that this paper may be of interest as a contribution to the study of micro-powder ferromagnetic materials, especially as there appears as yet to have been no published work on powders known to contain acicular particles (see Weil 1951, p. 444). An extension to the effects of wider departures from stoichiometric chemical composition, involving partial transformation to the antiferromagnetics α-Fe_2O_3 and FeO, is contemplated for future work.

ACKNOWLEDGMENTS

It is a pleasure to record my indebtedness to my colleagues Mr. P. T. Hobson and Mr. N. T. Winkworth for many results of experimental work and for the photographs from which the curves shown in the figure were constructed, and to Dr. Joy P. Stern for first introducing me to Wells' work on oxide structures and for many helpful discussions. My thanks are also due to Dr. K. Hoselitz for the benefit of his advice and criticisms after reading the original manuscript, and to the Directors of the Minnesota Mining and Manufacturing Company, Ltd., for permission to publish this paper.

REFERENCES

Bickford, L. R., Jr., 1950, *Phys. Rev.*, **78**, 449.
Birks, J. B., 1950, *Proc. Phys. Soc.* B, **63**, 65.
Bozorth, R. M., 1948, *Z. Phys.*, **124**, 519.
Chevallier, R., 1951, *J. Phys. Radium*, **12**, 172 (Colloque International de Grenoble, 1950).
Domenicali, C. A., 1950, *Phys. Rev.*, **78**, 458.
FIAT Report, 1947, *Final Report No.* 923 (London: H.M. Stationery Office).
Gorter, E. W., 1950, *Nature, Lond.*, **165**, 798.
Hobson, P. T., Chatt, E. S., and Osmond, W. P., 1947, *Electronic Engng.*, **19**, 383.
Kaye, G. W. C., and Laby, T. H., 1944, *Physical and Chemical Constants* (London: Longmans, Green), p. 134.
Michel, A., Chaudron, G., and Bénard, J., 1951, *J. Phys. Radium*, **12**, 189 (Colloque International de Grenoble, 1950).

Néel, L., 1947 a, *C.R. Acad. Sci., Paris*, **224**, 1550; 1947 b, *Ann. Univ. Grenoble*, **22**, 299; 1948 a, *Ann. Phys., Paris* (12), **3**, 137; 1948 b, *J. Phys. Radium*, **9**, 184; 1949, *Ann. Phys., Paris* (12), **4**, 249.

Stoner, E. C., 1934, *Magnetism and Matter* (London : Methuen), p. 529; 1950, *Rep. Prog. Phys.*, **13**, 83 (London : Physical Society).

Stoner, E. C., and Wohlfarth, E. P., 1947, *Nature, Lond.*, **160**, 650; 1948, *Phil. Trans. Roy. Soc.* A, **240**, 599.

Verwey, E. J. W., and Heilmann, E. L., 1947, *J. Chem. Phys.*, **15**, 174.

Weil, L., 1947, *C.R. Acad. Sci., Paris*, **225**, 229; 1948, *Ibid.*, **227**, 1327; 1951, *J. Phys. Radium*, **12**, 437 (Colloque International de Grenoble, 1950).

Weiss, P., and Forrer, R., 1929, *Ann. Phys., Paris*, **12**, 279.

Wells, A. F., 1945, *Structural Inorganic Chemistry* (Oxford : Clarendon Press), p. 334.

The Emission from Oxide-coated Cathodes in an Accelerating Field

By D. A. WRIGHT and J. WOODS
Research Laboratories, The General Electric Co. Ltd., Wembley

Communication from the Research Staff of the M.-O. Valve Co. at the G.E.C. Research Laboratories, Wembley, England; MS. received 17*th July* 1951, *and in amended form* 8*th October* 1951

ABSTRACT. It is shown that the steep rise of current with accelerating field in a diode with an oxide-coated cathode cannot be fully explained either by surface roughness or by patch effect. This leads to a theoretical treatment, based on the existence of a space-charge zone immediately inside the coating. The charge in this zone is shown to vary with applied field and with current density, and with certain coating parameters. The variation in the charge leads to a variation in work function, and thereby to a dependence of emission on field strength, which is to be combined with the normal Schottky effect. The effects of partial space charge on the relationship between field and anode voltage are discussed, and the diode characteristics of oxide cathodes can then be calculated for accelerating fields as a function of the coating parameters. The results are shown to be in agreement with experiment, and to confirm values of coating conductivity, electron density and mobility as determined directly by conductivity and Hall effect measurements.

§ 1. INTRODUCTION

The thermionic emission from a cathode with an electron accelerating field at the cathode surface was first treated theoretically by Schottky (1914). The relationship between emission current and applied field which he deduced has been verified experimentally for emitters such as clean tungsten. For emitters such as thoriated tungsten, the Schottky law is found to apply only at very high field strength, and the departure at lower fields has been attributed (e.g. Becker 1935) to a patchy variation in work function over the cathode surface. Much larger deviations from the Schottky law are commonly found in the case of oxide-coated cathodes. These have been illustrated in previous publications (Coomes 1946, Danforth and Goldwater 1949, Sproull 1945, Violet and Riethmuller 1949, Wright 1949 a) and have the following characteristics: (i) At the higher operating temperatures, that is, above 1050° K, the gradient of the curve beyond the 'knee' may be so steep that the knee is only just detectable as a small change in curvature. The steepness near the knee is in part due to the fact that a partial space charge remains present when the field at the cathode has small positive values. This is discussed in § 6. (ii) There is frequently an inflection at higher voltages, where the gradient increases. Both of these observations refer to pulsed operation. (iii) The gradient beyond the knee is steeper than the Schottky law would predict, over a wide voltage range at temperatures so low that the partial space-charge effect is confined to voltages near the knee. Observations of this type are normally made in d.c. operation at 500–700° K.

Fig. 1 illustrates in the curves A, the low temperature d.c. behaviour in a plane-parallel diode for a normal sprayed coating. Fig. 2 shows curves obtained at higher temperatures in pulsed operation. These curves have been obtained in the course of the present experimental work. The low temperature plots commonly have slopes from two to four times as great as the theoretical Schottky slopes.

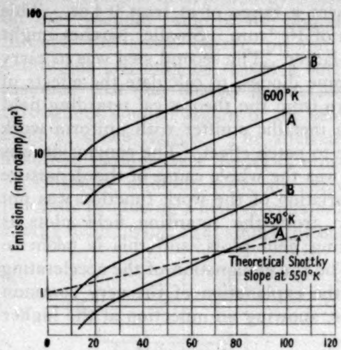


Fig. 1. Accelerating field plots at low temperature.

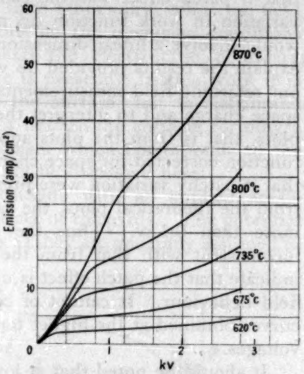


Fig. 2. Pulsed diode characteristics. (Diode with 0·7 mm anode–cathode space.)

It has been suggested that the observed curves are the consequence of the roughness of the surface of a sprayed coating, combined with the variation in work function over the surface (Hung 1950, Arizumi and Narita 1951). It has also been pointed out by Morgulis (1947) that when the electron emitter is a semiconductor, field penetration of the surface should enhance the electron density near the boundary, so raising the emission above the value predicted by Schottky's law.

The purpose of the present paper is threefold: (*a*) To show that roughness and patch effect are not adequate explanations of the curves in figs. 1 and 2. (*b*) To show that in broad terms Morgulis' theory of field penetration is able to predict these curves, though in detail an important correction to his equations is necessary. (*c*) To combine the theory of field penetration with the treatment of the case of partial space charge, thus enabling diode characteristics to be plotted as a function of anode voltage in the accelerating field region. These can then be compared with curves obtained in practice, such as those in fig. 2.

§2. EXPERIMENT ON DIODE CHARACTERISTICS

In order to test the importance of roughness, a coating was sprayed on a flat nickel surface and was then compressed mechanically, using a hardened steel surface polished to an optical flat, and a pressure of 2 tons/in^2. This gave a much denser packing of particles than is typical of sprayed coatings, and a surface with a high degree of smoothness. It had a glazed appearance, with a considerable polish. Diode tests with these smooth coatings gave the curves B in fig. 1, with the same slopes as curves A. These results, typical of those for a number of diodes, show that the macroscopic roughness of a sprayed coating is not an important source of the difference between such a coating and a metallic surface.

The patch theory was checked by comparing the behaviour of diodes in accelerating and in retarding fields. Curves such as those in fig. 1 were analysed in terms of Becker's work on the patch effect (Becker 1935), and it was concluded that if patch effect was the cause of the departure from Schottky's law, the variation in work function $\Delta\phi$ must be over a range of at least 0·4 ev. This would involve a linear dimension of patch of 10^{-4} mm. Smaller patches might explain the results provided $\Delta\phi$ were even larger. The second step was to carry out retarding field measurements in the same diodes, to calculate the effects of space charge and to interpret the departure from the theoretical retarding field plots, that is from the plots applying to a metallic emitter with uniform work function corrected for space charge, as due to patch effect. The conclusion was that if patchy variation were present, and was the whole cause of the departure from the theoretical plots, the range of variation of the work function was not more than 0·1 ev. Thus the conclusion from the retarding field plots is inconsistent with that from the accelerating field plots, and this is taken to indicate that the patch effect is unlikely to be the explanation of the accelerating field behaviour. It cannot of course be the explanation of the very common curves obtained at the higher temperatures, showing an inflection at the higher voltages.

It should be noted that at low temperature, the coating resistance may well be large enough to affect the retarding field characteristic. In this case patch effect, if present, would not be the whole cause of the deviation from the theoretical plot applying for a metallic emitter, and the range of variation of work function would then be even less than 0·1 ev.

The experimental results of this section justify a theoretical consideration of the effect of the applied field on the energy levels near the surface of an oxide coating. The treatment is based on the fact that when current flows from one dielectric medium to another there is in general in the steady state a charge in the region of the boundary between the media. This charge varies with current density and with applied field. The following sections give the theory of this effect when the oxide cathode is considered as a homogeneous dielectric medium, in which the charge exists in a space-charge zone immediately inside the coating.

An alternative treatment appears to be possible in which the charge is taken to be localized as a plane sheet over the surface of the dielectric. This would need separate consideration. It should be noted that the treatment given below, while referring specifically to the oxide cathode, has general application to all cases where current flows from a dielectric medium to a vacuum, whatever the mechanism of electron emission.

A theory of this type has already been presented by Morgulis (1947), but his paper does not give the derivation of the expression finally presented, and the present paper shows that Morgulis' solution was restricted to a special case of a more general solution, since he omitted the effect of the finite resistance of the coating. This would lead to serious errors under typical practical conditions, especially at low temperatures. The general solution is given below, and diode characteristics are determined for the combined effects of applied field on the space-charge zone in the coating and on the external potential barrier.

§ 3. SPACE-CHARGE ZONE AT THE COATING SURFACE

Since an oxide cathode can be regarded as a dielectric medium with a finite resistance, the usual electrostatic boundary conditions must be obeyed at the oxide–'vacuum' interface. When a positive anode potential is applied to a diode with an oxide cathode, a current flows, and the field strength in the bulk of the coating is determined by considerations from Ohm's law, while the field on the 'vacuum' side of the boundary depends on the operating conditions through the space charge present. In general the electrostatic boundary conditions can only be met if a surface charge is present. In this work it is assumed that the charge is not localized at the surface, but is distributed throughout a small region of the semiconductor next to the surface, forming a layer of space charge. The alternative mechanism which corresponds to the adsorbed barium layer theory of emission will not be discussed here.

Consider a diode with an oxide-coated cathode with an applied anode potential producing a field X volts/cm at the surface of the oxide. The potential energy diagram fig. 3 (*a*) shows the energy levels as they are generally assumed to

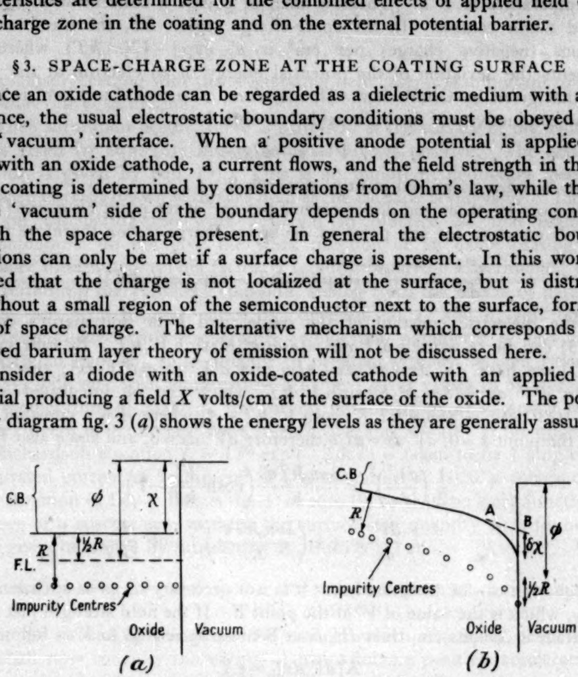


Fig. 3. Energy levels in an oxide coating: (*a*) at coating–vacuum interface in absence of a field; work function $=\phi$; (*b*) at the coating–vacuum interface as modified by positive field at surface (positive means a field accelerating electrons to the anode); work function $=\phi-\delta\chi$.

be. This will apply for the case where no space-charge zone is present at the coating surface. The work function ϕ is given by

$$\phi=\tfrac{1}{2}R+\chi \qquad \ldots\ldots(1)$$

where R is the activation energy, and χ the electron affinity of the coating. Fig. 3 (*b*), however, shows the modification of this diagram, necessitated by the considerations outlined above, when an applied field X is present and a current I flows. The energy levels are tilted in the bulk of the oxide by an amount determined by Ohm's law, while they are curved in the space-charge region and in consequence produce a deviation $\delta\chi$, from the position they would have occupied had no space charge been present, and had Ohm's law been considered alone. $\delta\chi$ is the change in work function ϕ as given by eqn. (1), which it is required to calculate. In the case of fig. 3 (*b*), X is a positive field and the situation represents a decrease of work function due to a negative space charge.

Referring to fig. 3 (*b*), let the bottom of the conduction band in the bulk oxide, at point A say, be taken as the zero of potential energy V, and let the concentration of free electrons in the conduction band at this point be n_0, if the temperature is T° K. There is then of course a concentration n_0 also of ionized donor centres. In the space-charge region AB, the number of free electrons (negative charge) per cm^3 is $n_0 \exp\{-V'(x)/\boldsymbol{k}T\}$ where $V'(x)$ represents the deviation of the potential energy of an electron at the point x from that given by Ohm's law. If, therefore, a current I amp/cm^2 is flowing, and the conductivity of the bulk oxide is σ ohm^{-1} cm^{-1}

$$V' = V - \boldsymbol{e}Ix/\sigma. \qquad \ldots\ldots(2)$$

$x=0$ at A and is measured positive towards B.

The number of ionized donor centres per cm^3 at the point x is $n_0 \exp(V'/\boldsymbol{k}T)$. Poisson's equation for any point in the space-charge region therefore becomes

$$\frac{d^2V}{dx^2} = \frac{4\pi n_0 \boldsymbol{e}^2}{K}\left[\exp(V'/\boldsymbol{k}T) - \exp(-V'/\boldsymbol{k}T)\right]. \qquad \ldots\ldots(3)$$

K is the dielectric constant of the oxide, and $\boldsymbol{k}$ is Boltzmann's constant. Eqn. (3) can be re-written $d^2V/dx^2 = (8\pi n_0 \boldsymbol{e}^2/K)\sinh V'/\boldsymbol{k}T$. By differentiating eqn. (2) we have $dV'/dx = dV/dx - \boldsymbol{e}I/\sigma$ and $d^2V'/dx^2 = d^2V/dx^2$; therefore we can put $d^2V'/dx^2 = (8\pi n_0 \boldsymbol{e}^2/K)\sinh V'/\boldsymbol{k}T$, which can be integrated at once, giving $\frac{1}{2}(dV'/dx)^2 = (8\pi n_0 \boldsymbol{e}^2 \boldsymbol{k}T/K)\cosh V'/\boldsymbol{k}T + A$, where A is a constant.

At the point $x=0$, $dV/dx = \boldsymbol{e}I/\sigma$, therefore $dV'/dx = 0$, and since also $V'=0$

$$\frac{1}{2}\left(\frac{dV'}{dx}\right)^2 = \frac{8\pi n_0 \boldsymbol{k}T\boldsymbol{e}^2}{K}\left(\cosh\frac{V'}{\boldsymbol{k}T} - 1\right)$$

so that

$$\frac{dV'}{dx} = \pm 4\left(\frac{2\pi n_0 \boldsymbol{k}T\boldsymbol{e}^2}{K}\right)^{1/2}\sinh\frac{V'}{2\boldsymbol{k}T}. \qquad \ldots\ldots(4)$$

This equation can be integrated, but it is not necessary to go to such lengths to find $\delta\chi$, which is the value of V' at the point B. If the field strength just outside the surface is X volts/cm, then dV/dx at B must be related to X as follows:

$$K(dV/dx)_{\mathrm{B}} = \boldsymbol{e}X. \qquad \ldots\ldots(5)$$

Combining this with the expression for dV'/dx we have

$$K\left(\frac{dV'}{dx}\right)_{\mathrm{B}} + \frac{K\boldsymbol{e}I}{\sigma} = \boldsymbol{e}X. \qquad \ldots\ldots(6)$$

Inserting this value for $(dV'/dx)_{\mathrm{B}}$ into eqn. (4) we obtain an expression containing $\delta\chi$, namely

$$\frac{\boldsymbol{e}X}{K} - \frac{\boldsymbol{e}I}{\sigma} = \pm 4\left(\frac{2\pi n_0 \boldsymbol{k}T\boldsymbol{e}^2}{K}\right)^{1/2}\sinh\frac{\delta\chi}{2\boldsymbol{k}T}.$$

If the left-hand side of this equation is positive, $\delta\chi$ represents a decrease in ϕ, and in consequence the negative sign should be taken, and

$$\frac{X}{K} - \frac{I}{\sigma} = -4\left(\frac{2\pi n_0 \boldsymbol{k}T}{K}\right)^{1/2}\sinh\frac{\delta\chi}{2\boldsymbol{k}T}. \qquad \ldots\ldots(7)$$

At saturation when $X=0$, eqn. (7) becomes

$$\frac{I_0}{\sigma}=4\left(\frac{2\pi n_0\boldsymbol{k}T}{K}\right)^{1/2}\sinh\frac{\delta\chi_0}{2\boldsymbol{k}T} \qquad \ldots\ldots(8)$$

where $\delta\chi_0$ is now an increase in work function, over that defined by eqn. (1), and I_0 is the saturation current. The work function ϕ' at saturation should be

$$\phi'=\phi+\delta\chi_0=\tfrac{1}{2}R+\chi+\delta\chi_0. \qquad \ldots\ldots(9)$$

§ 4. THE WORK FUNCTION AT ZERO FIELD

Before proceeding with the accelerating field case, it is important to consider the behaviour at saturation.

When saturation occurs, there is zero field at the surface of the cathode and $\delta\chi_0$ is given by eqn. (8); this quantity is always positive. The actual work function is given by eqn. (9). This is an important deduction, because the saturated current drawn with zero field at the cathode is the value of the current used in calculating the work function in practice, so that experimental determinations do not give the work function of an oxide coating as it is normally interpreted. This factor is in addition to the experimental difficulty of determining the zero-field emission, arising from the effects of partial space charge, see § 6. Further, since $\delta\chi_0$ depends on I_0 and σ, and both these quantities change with temperature, while this parameter is also contained explicitly in eqn. (8), it is difficult to predict, without detailed examination, how $\delta\chi_0$ will vary with temperature. Table 1 gives values of $\delta\chi_0$ for a coating at a temperature of 1000° K, for various combinations of n_0 and σ, assuming that a current of 9 amp/cm² flows in each case and $K=10$. If the quantity A in the Richardson equation $I_0=AT^2\exp(-\phi/\boldsymbol{k}T)$ is taken to be 1 amp/cm² deg², a saturated current of 9 amp/cm² at a temperature of 1000° K corresponds to a work function of 1 ev. This is the ϕ' of eqn. (9) so that the work function in the absence of a space-charge zone at the surface (the quantity ϕ) is found for the cases given in table 1 by subtracting $\delta\chi_0$ from $\phi'=1$ ev.

5. REDUCTION OF THE WORK FUNCTION IN THE ACCELERATING FIELD REGION

In the preceding paragraph the case in which $X=0$ has been discussed. We shall now consider the effects of maintaining a positive (accelerating) field at the cathode surface. Eqn. (7) is the starting point, giving an expression from which $\delta\chi$ may be calculated. This equation cannot be solved in general terms, as will be discussed later, but a useful approximation is to assume that the resistance of the coating is so low that the term I/σ can be neglected in comparison with X/K. Eqn. (7) then becomes, after rearranging,

$$\delta\chi=-2\boldsymbol{k}T\sinh^{-1}\frac{X}{4(2K\pi n_0\boldsymbol{k}T)^{1/2}} \text{ and } \delta\chi_0=0. \qquad \ldots\ldots(10)$$

This is the result obtained by Morgulis and is restricted to the case where this assumption is justified. Fitting numerical values to eqn. (7) shows that the approximation is justifiable only when $\sigma \geqslant 10^{-1}$ ohm^{-1} cm^{-1}, when $I_0=9$ amp/cm².

Using this simplification, the change in work function at zero field is zero. Thus, for all coatings where $\sigma \geqslant 10^{-1}$, the value of $\delta\chi_0$ (§ 4) is negligible. This situation corresponds to the case where the energy levels, fig. 3 (*b*), in the bulk of the coating are horizontal instead of sloping to the right. Eqn. (10) is easily solved, and values of $\delta\chi$ are given in table 2 for various field strengths on the assumption that $K=10$, $T=1000°$ K, and $n_0=10^{14}$ per cm^3. A discussion of probable values of n_0 is given in §9. In fig. 4, curve A shows the variation of current with field strength making these same assumptions, and with a diode such that saturation occurs at 9·0 amp/cm^2. Curve B is the corresponding case with the temperature reduced to 900° K when the saturated current is 2·0 amp/cm^2. This means a lower value of n_0 and a higher resistance of the coating. The values are deduced taking an activation energy $R=1{\cdot}4$ ev. Curves C and D are the Schottky lines for the same diode at the same temperature. Since these two effects occur simultaneously, they must be combined in order to give the total current flowing. This is shown by curve E for the higher temperature of 1000° K. Fig. 5 shows the energy level diagram when the Schottky effect is taken into account as well. It should be noted here that the image force expression, used in calculating the Schottky effect, is different for a semiconductor from that for a metal. The difference is however negligible when the dielectric constant is high, as it is for barium oxide.

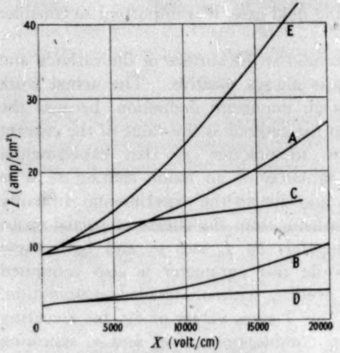


Fig. 4. Current density as a function of applied field, due to Schottky effect and to effect of field on surface space charge. Coating with $n_0=10^{14}$ per cm^3.
A. Effect of surface charges at 1000° K.
B. Effect of surface charges at 900° K.
C. Schottky effect at 1000° K.
D. Schottky effect at 900° K.
E. Combined effect at 1000° K.

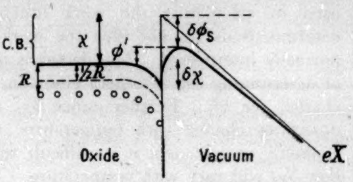


Fig. 5. Energy levels in an oxide coating when space-charge and Schottky effects are both taken into account, and volt drop in coating is neglected. In absence of field, work function is $\phi=\frac{1}{2}R+\chi$. With applied field X, work function is $\phi=\phi'-\delta\chi-\delta\phi_s$.

The approximate method cannot be used for coatings with σ less than 10^{-1} ohm^{-1} cm^{-1} because the potential drop in the coating cannot be ignored. In this general case the work function at zero field is $\phi'=\phi+\delta\chi_0$ where $\delta\chi_0$ is given by eqn. (8), and the first step is to solve this equation. It is supposed that, in all cases to be considered, ϕ' remains constant, with the value of 1 ev, corresponding to an emission of 9 amp/cm^2 at 1000° K, and 2 amp/cm^2 at 900° K. These are the observed emissions for a coating in a very good state of activation. When different values of n_0 or σ are selected for consideration, the constancy of ϕ' implies that corresponding values of ϕ are selected to compensate for the change in χ_0, as in the previous section. With chosen values of n_0 and σ, $\delta\chi_0$ and therefore ϕ' will vary with temperature. This effect is however small, and has been neglected in obtaining solutions at 900° and 1000° K.

Having found $\delta\chi_0$, the procedure is as follows. Choose some value of $I > I_0$, and find the total reduction of work function $\delta\phi'$ required to produce this current. The value of $\delta\chi$ which would be necessary to effect this reduction, if the Schottky lowering is ignored, is $\delta\chi_0 - \delta\phi'$. Insert this value into eqn. (7) and obtain a solution for X. Now due allowance for the Schottky term must be made, so with this value of X, $\delta\phi_S$, the change in work function due to this effect, is determined by the equation

$$\delta\phi_S = -\epsilon^{3/2} X^{1/2}. \qquad \ldots\ldots(11)$$

This requires a modification of the first estimate of $\delta\chi$ since $\delta\phi'$ is constant and given by

$$\delta\phi' = \delta\chi + \delta\phi_S - \delta\chi_0. \qquad \ldots\ldots(12)$$

(The quantities $\delta\phi'$, $\delta\chi$ and $\delta\phi_S$ which represent decreases of work function are negative, while $\delta\chi_0$ is positive.)

Inserting the new value of $\delta\chi$ into eqn. (7) the whole procedure need only be repeated three or four times before a solution for X accurate to 1 part in 200 is obtained. In this way curves of I as a function X can be drawn when the two effects are considered together.

§6. THE EFFECT OF PARTIAL SPACE CHARGE

In order to compare the results with experiment, it is desirable to plot the emission current in an accelerating field as a function of anode voltage rather than of applied field. It is known that the field is given by the anode voltage V_a divided by the anode–cathode spacing a when the voltage is very large, but in the region near the 'knee' this is by no means the case, since although there is no space-charge minimum of potential in the cathode–anode space above the knee, there is nevertheless some space-charge effect to be considered. The field is considerably less than V_a/a for voltages just beyond the knee in the diode characteristic. It is to be expected that at high current density the range of voltage over which the field is less than V_a/a will be large, and consequently that a considerable range of the characteristics in fig. 2 will be affected by these considerations.

The treatment of this case of partial space charge has been given by several writers, in particular by Ivey (1949). It is not possible for the present purpose to use Ivey's results directly, but the initial procedure is the same, and Ivey's solution will be used as the starting point. It should be noted that this involves the simplifying assumption that the electrons are emitted with zero velocity.

Consider a planar diode with a clearance of a cm between anode and cathode, and let a potential V_a be applied to the anode so that a current I amp/cm^2 flows and an accelerating field X v/cm appears at the cathode surface. By integrating Poisson's equation for the partial space charge in the anode–cathode space, and inserting the boundary conditions appropriate to the problem, the following relation between V_a, I and X can be obtained:

$$(v_a + L)^{3/2} - 3L(v_a + L)^{1/2} + 2L^{3/2} = 3a(8\pi \boldsymbol{m}\boldsymbol{e}I)^{1/2}/2\boldsymbol{m} \qquad \ldots\ldots(13)$$

where $v_a(2\boldsymbol{e}V_a/\boldsymbol{m})^{1/2}$ and $L = \boldsymbol{e}X^2/8\pi\boldsymbol{m}I$. $\boldsymbol{e}$ is the charge and $\boldsymbol{m}$ the mass of an electron. This is essentially the result obtained by Ivey and others, expressed

.B
5
2

in a form more suited to the present requirements. In using this equation, the procedure is to deal with a particular relationship between the current I in an accelerating field, the field strength X, and the emission at zero field I_0, and so to derive values of I as a function of X for particular values of I_0. This can be done for example for the Schottky law, or for the simplified relation (10), or for the full expression (7) after solution by the method of § 5.

Calculations have been made first for two particular cases of cathodes operating at high current density, but where it is assumed that the Schottky law applies. This is of interest since it shows that the field is less than V_a/a over a considerable voltage range, causing an apparent departure from the Schottky law if it is mistakenly supposed that X and V_a/a are identical.

In the first instance, assuming an electrode separation of 1 mm, and a thoriated tungsten cathode with work function equal to 2·63 ev, and $A=5$ amp/cm² deg² the saturated emission I_0 is 4·58 amp/cm² if $T=2000°$ K. Using this value in the equation for Schottky's law $I=I_0 \exp(\boldsymbol{e}^{3/2}X^{1/2}/\boldsymbol{k}T)$, the current in the accelerating field region was calculated for various values of X. The straight line A which results when log I is plotted against $X^{1/2}$ is shown in fig. 6. Making use of the analysis above, the values of V_a necessary to produce the fields X at the particular current densities in question have been determined. Curve B, fig. 6, shows log I plotted against $(V_a/a)^{1/2}$. Curves C and D are similar

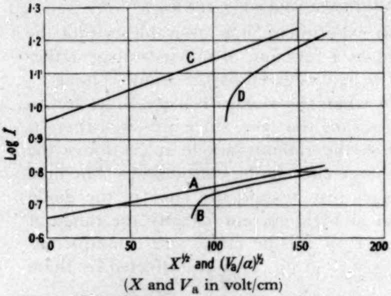


Fig. 6. Logarithmic plot of diode current in an accelerating field as function of field strength and of anode voltage. The true and apparent Schottky lines, A and B, are for a thoriated tungsten cathode and C and D for a mixed oxide cathode.

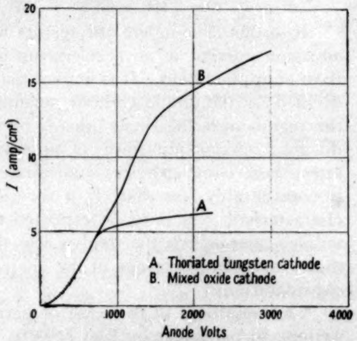


Fig. 7. Complete diode characteristics under the combined influence of space-charge and of Schottky effect.

plots for a diode with the same separation but with a mixed oxide cathode with $\phi=1$ ev and $A=1$ amp/cm² deg², maintained at a temperature of 1000° K so that $I_0=9$ amp/cm².

It will be observed that the lower curve of both sets tends to the true Schottky line at high anode potentials: this is due to the gradual elimination of space charge. The slopes in the accelerating field region are obviously in excess of the true Schottky slopes, but the ratio where the lower curve is almost linear is only 1·5 in the mixed oxide case and 1·3 in that of the thoriated tungsten. These figures apply to the portion of each plot shown, namely up to a field strength of 2×10^4 volt/cm, which in both diodes is obtained with an anode potential three times larger than that required to produce saturation. At higher voltages still the disparity between the true and apparent Schottky lines will be less.

Complete diode characteristics are drawn in fig. 7; graph A is for thoriated tungsten, and graph B for the mixed oxide cathode. An interesting feature is revealed by these curves. Reference to curve B shows that the apparent saturated, or 'knee', value of emission is 12·5 amp/cm^2, whereas the true saturated emission is in fact 9 amp/cm^2. Similarly the apparent value from curve A is 5·0 amp/cm^2, whereas the true value of I_0 is 4·58 amp/cm^2. The deviation from the space-charge line is so small in curve B between the value of the zero field saturated emission, 9 amp/cm^2, and the apparent 'knee' value at 12·5 amp/cm^2 that it will be very easy to be misled in examining experimental plots, and to assume that the emission is considerably greater than its true zero-field value. This is of considerable importance, as it is customary to define the maximum space-charge-limited current as the true emission. This small deviation is of course due to the fact that the field rises very steeply with anode voltage when the field first becomes positive at the cathode surfaces. It will be noted that the reverse situation obtains if a plot is made of log I against $(V_a/a)^{1/2}$ as in fig. 6, curve D, and if the linear part of this plot is extrapolated back to zero V_a. This will indicate an emission considerably less than the zero-field value. This arises from the fact that the values of V_a/a in curve D are much too small to remove partial space charge when the current density is so high. These results emphasize the difficulties in determining emission when its value is large, and indicate one reason why work functions determined in temperature ranges giving high emission differ from those obtained at lower temperatures.

§7. DIODE CHARACTERISTICS

The case to consider now is that of §5, where a set of curves has been obtained relating I with X, taking account of the Schottky law and of the field penetration effect. Graphical solution of eqn. (13) has been carried out for this case just as for the case where the Schottky law only is involved, again assuming a cathode–anode spacing of 1 mm. In fig. 8 these curves have been joined to the space-charge line and complete characteristics drawn. Curve A in fig. 8 corresponds to curve E in fig. 4, for a coating with an electron density n_0 of $10^{14}/\text{cm}^3$ and with $\sigma = 10^{-1}$ or more at 1 000° K, while curve B is for the same n_0, but gives the full solution of eqn. (7) where $\sigma = 10^{-2}$. The curve E in fig. 8 applies for the case where $n_0 = 10^{15}$ and $\sigma = 10^{-1}$ or more at 1 000° K, while the curves C, D and F correspond to A, B and E, but are plotted for a temperature of 900° K where I_0 is 2 amp/cm^2. Here the values of n_0 and σ are less than at 1 000° K by a factor which can be deduced, taking the slope of the curve of log σ or log n plotted against $1/T$ to be 0·7 ev. It is seen that the curves for accelerating fields are flatter the higher the value of n_0, and the lower the value of σ.

It is now necessary to consider the values of n_0 and σ which are likely to be encountered in practice. In table 1 values are chosen arbitrarily to show the effect on the work function, while in table 2, in fig. 4, and in curves A, B, C, D of fig. 8, the value of 10^{14} was chosen for n_0. The reason for this, and for considering also the case of $n_0 = 10^{15}$ in fig. 8, will appear from the following section.

Table 1. Increase in Work Function $\delta\chi_0$ at Saturation for various Values of n_0 and σ, for a saturation current of 9 amp/cm² at a temperature of 1000°K.

n_0 (free electrons/cm³)	10^{14}	10^{14}	10^{14}	10^{15}	10^{15}	10^{15}
σ (ohm⁻¹ cm⁻¹)	10^{-1}	10^{-2}	10^{-3}	10^{-1}	10^{-2}	10^{-3}
$\delta\chi_0$ (ev)	0·004	0·04	0·29	0·001	0·013	0·12

Table 2. Decrease in Work Function $\delta\chi$ for various Field Strengths X for a coating with low resistance and $n_0 = 10^{14}$ per cm³ at a temperature of 1000° K.

X (v/cm)	0	2500	5000	7500	11000	15000	20000
$\delta\chi$ (ev)	0	0·007	0·019	0·036	0·0525	0·07	0·093

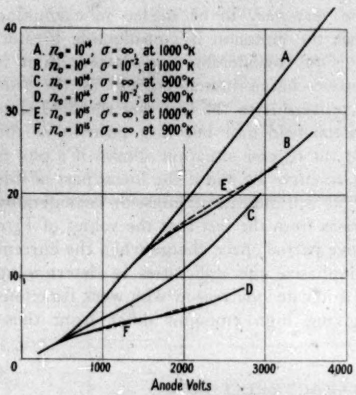


Fig. 8. Complete diode characteristics under the combined influence of space charge, Schottky effect and the field effect on the surface space-charge zone.

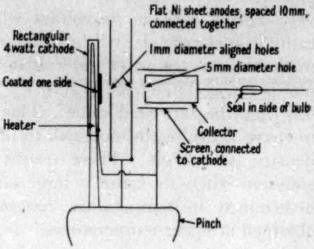

Flat Ni sheet anodes, spaced 10mm, connected together
Rectangular 4 watt cathode
1mm diameter aligned holes
5mm diameter hole
Coated one side
Seal in side of bulb
Heater
Collector
Screen, connected to cathode
Pinch

Fig. 9. Special tube with collector electrode.

§8. COATING RESISTANCE

Some measurements of coating resistance have been made, using a retarding field method. A tube was used as in fig. 9, in which a rectangular-section indirectly heated cathode is coated on one side with a known area of barium–strontium (or barium–strontium–calcium) carbonates. In most diodes tested, this was sprayed to a thickness of 0·1 mm, but tests were also made of the smooth compressed coating. An anode of two nickel plates has two aligned holes, centrally disposed with respect to the coating, and electrons emitted from the coating can pass through these holes to a shielded collector behind the anode. This is connected by means of a side connection through the glass bulb, minimizing leakage problems. At a given cathode temperature and anode voltage, there is a corresponding anode current I_a, and if the collector voltage is sufficiently high, a certain fraction kI_a flows through the holes to the collector. If however the collector potential is lowered below cathode potential, the current to the collector is in a retarding field, and falls logarithmically below the value kI_a. This gives a plot of collector current against collector voltage as in fig. 10, curve 1. More strictly, the collector current begins to fall when the collector potential becomes equal to V_1, the minimum potential near the cathode surface. When the emission is space-charge limited V_1 is less than the cathode surface potential.

If now, at constant cathode temperature, the anode voltage is altered to give an anode current I_a', the collector current at high collector voltage becomes kI_a', curve 2. The collector potential at which the collector current starts to fall (V_2) is the new value of potential at the potential minimum near the cathode surface. If the cathode were metallic, the potential at this point would vary in a known way (Langmuir and Compton 1931) with the value of I_a. With the oxide coating, a potential drop will exist across the coating, and will vary with the magnitude of I_a. The displacement, with I_a, of the collector voltage at which the collector current starts to fall therefore gives a method for investigating coating resistance.

Fig. 10 shows a family of curves, for different values of I_a, for collector current against collector voltage. Plotting the critical voltage against the anode current gives in fig. 11 a line whose slope would give the coating resistance if there were no change in the depth of the potential minimum below the potential at the coating outer surface. Correction for variation in this depth gives the dotted lines, the slope of which measures the coating resistance. Repeating at different temperatures we obtain the resistance values shown in fig. 12. It will be noted that the slope of the plot of log R against $1/T$ in the linear region is 0·7 ev, and that the conductivity (at 1000° K) is 3×10^{-3} ohm^{-1} cm^{-1}.

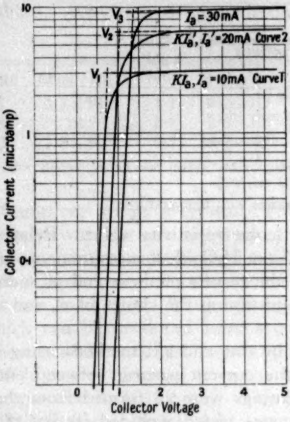


Fig. 10. Collector current variation.

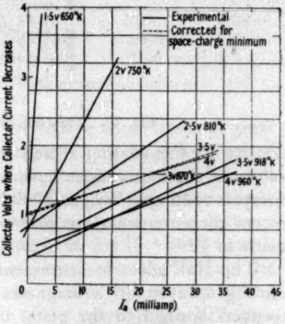


Fig. 11. Relation between anode current and critical collector voltage.

It will be noted that at temperatures near 1000° K, the slope of log R appears to be reduced. This is not observed with other methods of measurement, and may be due to inaccuracies introduced when the potential difference across the coating becomes similar to or less than the depth of the potential minimum outside the coating. There is however a further possibility introduced by the work of §4: there it is shown that with zero or retarding fields at the coating surface, there is a region of positive space charge immediately inside the coating. The presence of this zone will modify the potential variation through the coating and will make this method of estimating coating resistance inaccurate in cases where the potential difference across the coating is only a few tenths of a volt.

§ 9. COATING CONDUCTIVITY AND ELECTRON DENSITY

The values of conductivity derived from fig. 12 are in good agreement with those obtained for similar coatings by other methods (Wright 1947, 1950). It is of interest to note that the highly compressed type of coating (§ 2) gave resistance values similar to those of coatings as sprayed. In the latter case, the density of the coating is near unity, while in the former case the density of the same amount of coating, after compression, is near five. The similarity in the resistance value indicates that the resistance as measured for sprayed coatings is primarily that of the coating material, and is not largely affected either by phenomena in the voids, or by high resistance at particle contacts. It was necessary in § 7 to adopt values for coating resistivity and for the electron density in the coating which matches a particular resistivity value. Related values of these quantities have been obtained from Hall effect measurements on sprayed coatings (Wright 1949 b, 1950), and the results of these and of more recent measurements are summarized below, justifying the choice of n_0 and σ made in § 7.

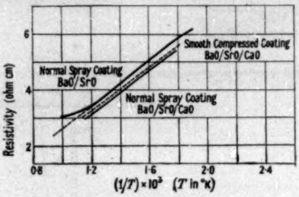


Fig. 12. Variation of resistivity of coatings with temperature.

The Hall effect measurements were made by d.c. and a.c. methods using a coating sprayed on a magnesia base plate, the current passing between end contacts applied to the plate. The oxide coatings were all formed from the carbonates in the usual way, and the carbonates tested included (i) $BaCO_3$, (ii) $(Ba\text{–}Sr)CO_3$ in equimolecular proportions prepared (*a*) by mixing, (*b*) by co-precipitation from the nitrate with sodium carbonate, (*c*) by co-precipitation with ammonium carbonate, and (iii) $(Ba\text{–}Sr\text{–}Ca)CO_3$ with molecular proportions 49 : 44 : 7 and 56 : 37 : 7 prepared by sodium and ammonium precipitation. The results were not very different for the different coatings, though the triple carbonates gave about twice the conductivity and emission of the double carbonates.

It was found with these coatings that operation drawing emission at low current density with the cathode at 1050–1100° K caused an increase in coating resistance, and introduced the p-type conductivity at low temperatures previously reported. Activation by drawing emission at higher density and temperature caused a decrease in resistance, and gave the state described in table 3. This represents a lower degree of activation than is obtainable in cathodes where the coating is sprayed on a nickel base in the usual way. It was possible however to obtain a state more nearly resembling this well activated state by heating the coating in ethane vapour. It will be noted from table 3 that taking the double oxide coatings, the electron density is roughly proportional to the conductivity, indicating a fairly constant mobility, near 2000 cm^2 sec^{-1} volt^{-1}. With the triple oxide coatings, the mobility was similar, and the electron density was typically about twice as great. Thus the conductivity in the best activated state encountered in this work is near 10^{-2} ohm^{-1} cm^{-1}, and associated with this, the electron density is near 3×10^{13}.

While this represents the best state of activation encountered in the Hall effect studies, and while the measurements with retarding fields gave conductivities with similar values, in neither case was the emission as high as is obtainable under the best conditions. If a diode is made using an anode of strip tungsten, which can be well outgassed at 2600° K, and if the cathode is outgassed in such a position that the deposits from the cathode do not form on the anode, it is possible to obtain a zero-field pulsed emission of more than 10 amp/cm^2 at 1000° K. In diodes with nickel or copper anodes figures from 2–5 amp/cm^2 are more typical. These emissions are two or three times as great as those with the ethane-activated coatings on magnesia. A large number of measurements of emission and conductivity have been compared, and it is found that at a particular temperature, emission and conductivity for different cathode are in a fixed ratio. Thus it is probable that the best activated oxide coatings which can be produced have a conductivity as high as 3×10^{-2} ohm^{-1} cm^{-1}. These will be coatings from triple carbonates as listed above. Taking the mobility value obtained from the Hall effect measurements, the corresponding electron density is 10^{14}/cm^3. This justifies the use of the value of $n_0 = 10^{14}$ for a fully activated coating in §7.

Table 3. Activation of Equimolecular BaO–SrO Coating on Magnesia at 1000° K.

State	σ (ohm^{-1}cm^{-1})	Slope of log σ (ev)	R (cm^3/c)	n_0 (electrons/cm^3)	Mean free path l_0 (cm)	Mobility u_0 (cm^2/volt sec)	(I_0 amp/cm^2)
1	$3{\cdot}3\times10^{-4}$	1·1	$9{\cdot}2\times10^{6}$	8×10^{11}	$3{\cdot}3\times10^{-5}$	2600	0·08
2	10^{-3}	0·8	$2{\cdot}5\times10^{6}$	3×10^{12}	$2{\cdot}8\times10^{-5}$	2000	0·3
3	$3{\cdot}8\times10^{-3}$	0·7	$6{\cdot}2\times10^{5}$	$1{\cdot}2\times10^{13}$	$2{\cdot}6\times10^{-5}$	2300	1·3

1. After decomposing carbonates and outgassing. 2. After activating by drawing emission at 1150° K. 3. After reduction with ethane.

It is clear that porosity effects in the sprayed coating could invalidate the estimation of mobility, and that in this case, with the observed values of conductivity, the true electron density would be higher than the above estimates. For this reason in §7 some calculations were made for comparison with an electron density of 10^{15}; in coatings not fully activated the results of table 3 will apply directly, so that it is also of interest to consider the behaviour in an accelerating field when n_0 is of the order 10^{13}, and σ between 10^{-3} and 10^{-2}.

§ 10. COMPARISON WITH EXPERIMENT

The values of n_0 and σ deduced in §9, $n_0 = 10^{14}$ and $\sigma = 3\times10^{-2}$, would correspond in fig. 8 with a curve intermediate between A and B. This would agree well with experiment, as is indicated by comparison with fig. 2, where the coating was not quite so well activated as the theoretical coating discussed in fig. 8. It is necessary however to consider what deviations would be produced by different values of n_0 and σ, particularly as there is some doubt about the reliability of Hall effect measurement as an indication of the value of n_0 and of the mobility u_0, as mentioned in § 9.

If n_0 were equal to 10^{15}, curve E would be predicted if $\sigma \geqslant 10^{-1}$, while if the mobility were less than 2000, the lower conductivity would lead to curves flatter than E. Higher values of n_0 would give still flatter curves. In practice curves flatter than E or B in fig. 8 are not obtained with well activated cathodes at 1000° K, and this can be taken both as a verification of the theory and as an indication that n_0 is not in fact greater than 10^{15} in such cathodes. Moreover this fact indicates that if n_0 is 10^{15}, σ is not less than 10^{-1}, so that the mobility is not less than 600 cm^2/volt sec. The practical curves do however permit that n_0 may be less than 10^{15}, with correspondingly lower σ. Thus as shown already, n_0 can be 10^{14} and σ can be in the range 10^{-1} to 10^{-2}, while if n_0 were equal to 10^{13}, a curve resembling B but inflecting upwards more steeply would be obtained with $\sigma = 10^{-3}$. With σ greater than 10^{-2}, the curves would be steeper than is observed in practice. Thus the practical curves show that n_0 is less than 10^{15}, and that the mobility is greater than 600 but less than 6000. These results support the Hall effect measurements, and indicate that they are not widely in error due to the effects of conductivity in the voids in the coating. This was also indicated by the observation that the voltage drop across the coating is the same before and after compression, § 8. The other indication of the practical curves is that steeply inflected curves correspond with values of n_0 low compared with 10^{15}.

It is of interest to notice that Morgulis, neglecting coating resistance, deduced from Sproull's results (Sproull 1945) that n_0 has a value near 10^{15}. This is the case where $\sigma \geqslant 10^{-1}$, but if resistance is not neglected, the value of n_0 is not unique. Moreover the value of n_0 falls as T decreases, so that the Morgulis solution (10) would predict a very large increase in emission with field at low temperatures. This is not observed, as shown in fig. 1, although some increase in steepness at low temperature is indicated here, and has been observed by Arizumi and Narita (1951). This is accounted for by the use of the correct solution (7). The importance of the term in (7) including conductivity is of course clear by comparison of curves A and B, or C and D, in fig. 8.

ACKNOWLEDGMENT

In conclusion, the authors desire to tender their acknowledgments to the M.-O. Valve Co. Ltd., on whose behalf the work described in this publication was carried out.

REFERENCES

ARIZUMI, T., and NARITA, S., 1951, *J. Phys. Soc. Japan*, **6**, 118.
BECKER, J. A., 1935, *Rev. Mod. Phys.*, **7**, 95.
COOMES, E. A., 1946, *J. Appl. Phys.*, **17**, 647.
DANFORTH, W. E., and GOLDWATER, D. L., 1949, *J. Appl. Phys.*, **20**, 163.
HUNG, C. S., 1950, *J. Appl. Phys.*, **21**, 37.
IVEY, H. F., 1949, *Phys. Rev.*, **76**, 554.
LANGMUIR, I., and COMPTON, K. T., 1931, *Rev. Mod. Phys.*, **3**, 241.
MORGULIS, N., 1947, *J. Phys. U.S.S.R.*, **11**, 67.
SCHOTTKY, W., 1914, *Phys. Z.*, **15**, 872.
SPROULL, R. L., 1945, *Phys. Rev.*, **67**, 166.
VIOLET, F., and RIETHMULLER, J., 1949, *Ann. Radioélectricité*, **4**, 184.
WRIGHT, D. A., 1947, *Proc. Roy. Soc.* A, **190**, 394 ; 1949 a, *Proc. Phys. Soc.* B, **62**, 188 ; 1949 b, *Nature, Lond.*, **164**, 714 ; 1950, *Brit. J. Appl. Phys.*, **1**, 150.

The Effect of Mercury on Selenium

By H. K. HENISCH and E. W. SAKER
Physics Department, University of Reading

Communicated by R. W. Ditchburn; *MS. received* 30*th August* 1951

ABSTRACT. It is shown that the interaction between crystalline selenium and liquid mercury or mercury vapour results in the formation of mercuric selenide. When produced in this form, this material is an excess semiconductor of high conductivity. The associated volume and surface diffusion have been investigated using radioactive ^{203}Hg. Mercury added to selenium *before* the crystallization process results in material of high resistivity. This is ascribed to a reduction of the positive hole mobility.

§ 1. INTRODUCTION

In view of the great industrial applications of semiconducting selenium in the manufacture of rectifiers and photocells, the effect of accidental or deliberately added impurities is a problem of considerable practical importance. Among the various impurities which are important from this point of view, mercury occupies a special place because of the rapidity with which it acts upon selenium and the drastic nature of its effects. The knowledge available at the beginning of this investigation can be summarized as follows:

(*a*) The first recorded observations concerning the effect of metallic mercury on the resistance of selenium specimens appear to be due to Moss (1877). Temporary contact with mercury was found to result in a pronounced decrease in resistance.

(*b*) Brown (1913) found that the resistance of selenium specimens crystallized in the presence of mercury vapour increased on exposure to light. This observation was later confirmed by Tisdale (1918). No clear distinction between bulk and contact phenomena was made in these experiments and the exact interpretation of these observations is therefore uncertain.

(*c*) Williams and Thompson (1941) have reported on the rapid and destructive influence of mercury vapour on selenium rectifiers. In the presence of mercury vapour the rectification efficiency falls until the forward and reverse resistances of the rectifier discs become almost equal.

(*d*) Henisch and François (1950, 1951) have reported that the thermo-electric effect decreases rapidly in the presence of mercury vapour. After prolonged exposure thermo-electric effects of reversed polarity were observed. On the other hand, considerable quantities of mercury (e.g. up to 1·6% by weight) could be added to the selenium before melting and crystallization without causing any significant change of thermo-electric properties.

The experiments described in the following sections were carried out in order to obtain further information on the phenomena involved and, in particular, to investigate the extent to which mercury can penetrate into the selenium specimens. The results show that there is penetration to considerable depths which may be due to diffusion through the lattice or along grain boundaries. The principal electrical effects are shown to be due to the formation of mercuric selenide at the surface. The presence of this distinct phase has been confirmed by x-ray analysis. There is also evidence, from experiments with radioactive mercury, of rapid surface diffusion.

§ 2. EXPERIMENTAL METHODS AND RESULTS

Effect of Mercury added before Crystallization

Measurements of bulk resistivity were carried out on a series of rectangular specimens (20 mm × 5 mm × 3 mm) to which current electrodes and potential probes of gold had been evaporated *in vacuo*. These specimens were prepared by adding varying amounts of mercury to amorphous selenium powder which was then melted. After subsequent solidification, the resulting glassy solid was crushed to a fine powder, moulded to the above shape in a press and converted to the crystalline modification by heat treatment for 30 minutes at 210° c. The average density of these specimens was 4·27, compared with the x-ray density of 4·82, and did not differ significantly as between the various groups of different mercury content. The table gives the results in summarized form and shows that there is a pronounced *increase* of resistivity with increasing mercury content.

Specimen group	a	e	d	f
Mercury content (% by weight)	0	0·008	0·158	1·58
Mean resistivity (ohm.cm)	469 ± 90	618 ± 70	2766 ± 300	22060 ± 2000

Effect of Mercury Vapour on the Resistance of Crystalline Specimens

Specimens of the type described above were kept in a closed vessel in the presence of mercury, first in air at atmospheric pressure and later at reduced pressures. Typical results for the various groups are shown in fig. 1. It will be seen that the effect of mercury is in every case more rapid at reduced air pressures. No consistent differences were observed as between specimens belonging to the different groups. By the successive removal of thin surface layers and repeated measurements of resistance it was possible to estimate the depth of penetration. For the specimens referred to in fig. 1 this depth was of the order of 5×10^{-3} cm at the end of the experiment. By comparison, the thickness of the barrier layer on a selenium rectifier is usually estimated to be considerably smaller, e.g. between 10^{-5} and 10^{-4} cm. In view of the speed of

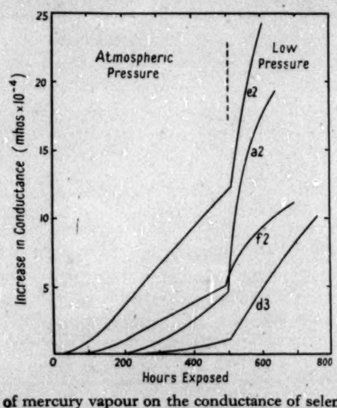


Fig. 1. Effect of mercury vapour on the conductance of selenium specimens.

the above diffusion process, such a layer could be rapidly penetrated in places which are not thoroughly protected by the counter electrode. Alternatively, the barrier layer could be short-circuited by mercury penetration near the edge of the counter electrode.

Action of Liquid Mercury on Selenium

Specimens of the type described above and some others of higher density (4·75), prepared by a casting technique, were kept immersed in liquid mercury at a temperature of 100° c for 100 hours. Sections were then made through the material and one of these is shown in fig. 2 (Plate). The appearance of the surface regions is visibly changed by this treatment, to a depth of about 0·25 mm. Accurate electrical measurements on the surface layer itself were not practicable, but it was confirmed that the layer has a high conductivity.

X-ray Tests

For purposes of identification and comparison, x-ray powder photographs were taken of normal crystalline selenium and also of two specimens which had been exposed to the influence of mercury in two different ways; fig. 3 (Plate) shows the results. The photograph A refers to a surface layer detached from a specimen similar to that shown in fig. 2. The lines are characteristic of a zinc-blende structure and lead to a lattice constant of 6·04 Å. The value quoted by Wyckoff (1931) for mercuric selenide is 6·07 Å. In view of the fact that all these specimens are expected to deviate considerably from stoichiometric composition, the two results are regarded as being in reasonably satisfactory agreement. Photograph B was obtained with a specimen of selenium which had been kept in the presence of mercury vapour in vacuum for 150 days at room temperature. It will be seen that lines corresponding to pure selenium and mercuric selenide are both represented. This shows partial conversion of the selenium into mercuric selenide.

Thermo-electric Tests

The specimen shown in fig. 2 was subjected to thermo-electric tests. It was clamped in a large electrode which was kept at room temperature, and a second contact of very small area was made by a wire which was heated electrically to about 60° c. The resulting thermo-electric force was then measured by means of a sensitive potentiometer circuit. The hot 'point' contact could be controlled by means of a micro-manipulator which made it possible to explore the thermo-electric properties of various surfaces. Results for measurements across one section of the specimen are shown in fig. 4. It will be seen that the thermo-electric effect changes sign at some distance from the exposed surface. The observed polarities correspond to n-type conduction in the surface region and to p-type in the interior. The magnitude of the thermo-electric power is considerably greater in the interior of the specimen than in the surface region. This is consistent with the previous observation that the surface region has a high conductivity. It is of interest to note that the electrical effects of mercury can be detected beyond the visible boundary of the selenide.

The Contact between Selenium and Mercuric Selenide

The above thermo-electric tests show the existence of a 'p–n contact' between the selenium and the mercuric selenide. The behaviour of such a contact is expected to differ somewhat from that of the more familiar p–n contacts in germanium, in view of the fact that the participating substances have not only a different conduction mechanism but also a different lattice. A simple experiment confirmed the existence of rectification in the expected direction. Fig. 5 shows the voltage–current relation of such a contact as obtained by oscillographic methods. Pronounced hysteresis effects were observed at 50 c/s.

Tests with Radioactive Mercury

In order to obtain more sensitive and direct evidence of mercury penetration some experiments were carried out with the radioactive isotope ^{203}Hg which has a convenient half-line of 43·5 days. The activity resulting from exposure to mercury vapour was measured with a Geiger counter and investigated as a function of depth. Fig. 6 shows typical results. In the dense specimen, B, the distribution follows approximately an exponential law of the form $N_x = N_0 \exp(-x/x_0)$, where N_0 and N_x denote the concentration of mercury at the surface and at a depth x respectively. For specimen B the length x_0 after 100 hours, exposure at room temperature and a pressure of 10^{-2} mm Hg was found to be 9×10^{-4} cm. In the amorphous specimen C it is smaller (e.g. 2×10^{-4} to 3×10^{-4} cm). Specimen A was very porous and a simple exponential distribution could not be expected. The term N_0 was found to be almost entirely determined by the roughness of the surface. It is smallest for surfaces which are highly polished. The depth of penetration measured for crystalline specimens is in agreement with the values estimated from electrical measurements.

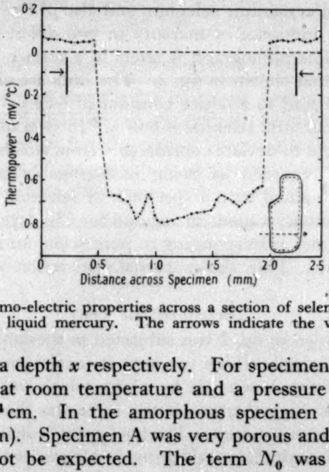


Fig. 4. Variation of thermo-electric properties across a section of selenium specimen exposed to the influence of liquid mercury. The arrows indicate the visible boundary.

To test for surface diffusion a small drop of mercury (2 mm in diameter) was placed on a disc of crystalline selenium and kept in a small enclosed vessel for a period of 40 hours. An autoradiograph was then taken and is shown in fig. 7. It will be seen that there is considerable diffusion along the surface. The mosaic structure is not due to the boundaries of individual ordered domains but arises from cracks in the material which are on a much larger scale.

§ 3. DISCUSSION AND CONCLUSIONS

The experiments confirm that the effect of mercury on selenium is not merely a surface phenomenon but is concerned with layers of considerable thickness. When a selenium specimen comes into contact with mercury or its vapour, mercuric selenide is formed on the outside, not only at high temperature as is well known (Mellor 1923), but also at room temperature, though more slowly. When formed under these conditions mercuric selenide is a good conductor of the excess type. The presence of even a very thin film of such a highly conducting n-type material is sufficient to account for the observed decrease in resistance, the reversal of the thermo-electric power already reported, and the deterioration of selenium rectifiers. The experiments, most of which were carried out on micro-crystalline specimens, could not distinguish between penetration of the lattice and diffusion along grain boundaries. The latter mechanism is perhaps more probable in view of the rapidity of the observed surface diffusion effects. The speed of the diffusion process is comparable with that of the thallium diffusion investigated by Gudden and Lehovec (1946).

When mercury is added to a selenium melt, it is able to take part in the subsequent crystallization process and could thus affect the length of the selenium chains (see, for instance, Krebs 1951). Its influence on the electrical properties of the material would thus be quite different from that of mercury added by diffusion after crystallization. The results of the resistivity measurements quoted in the table, together with the evidence derived from previous thermo-electric tests (Henisch and François 1950), indicate that mercury added before crystallization affects the mean free path of the carriers but not their concentration.

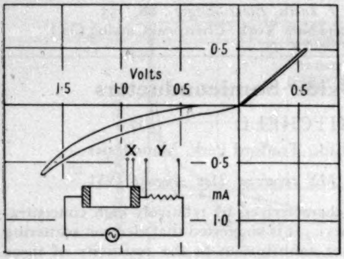


Fig. 5. Voltage—current relation of a contact between selenium and mercuric selenide.

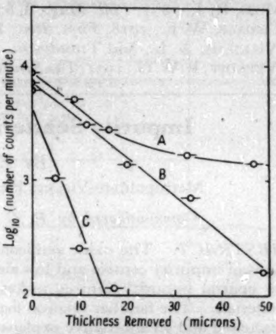


Fig. 6. Mercury concentration (proportional to number of counts) as a function of depth.

It seems desirable that experiments of the type here described should also be carried out on single crystals of selenium. The detailed mechanism of the interaction could also be usefully investigated by micro-balance techniques of the type successfully used in the study of oxidation processes. Lastly, as far as is known, mercuric selenide has not yet been systematically investigated for electrical properties and, in view of the interesting properties of other selenides, such an investigation is thought desirable.*

---

*A publication by A. I. Blum and A. R. Regel on the "Electrical Properties of Solid Solutions of Mercury-selenide and Selenium" (*J. Tech. Phys., USSR.*, 1951, **21**, 316) has come to the authors' notice since the preparation of the present paper.

§ 4. ACKNOWLEDGMENTS

The authors wish to thank Professor R. W. Ditchburn for placing research facilities at their disposal, Messrs. Standard Telecommunication Laboratories Ltd. for a research grant to one of them (E. W. S.), Mr. T. R. Scott of Standard Telecommunication Laboratories for his personal interest in the work and for the supply of selenium, and Dr. D. H. Tomlin and Mr. P. G. Hambling for their kind assistance and advice in connection with the radioactive tracer and x-ray experiments respectively.

REFERENCES

Brown, F. C., 1913, *Phys. Rev.*, **1**, 237 and **2**, 153.
Gudden, B., and Lehovec, K., 1946, *Z. Naturforsch.*, **1**, 508.
Henisch, H. K., and François, M. L. J., 1950, *Nature, Lond.*, **165**, 891; 1951, *Semi-conducting Materials*, ed. H. K. Henisch (London: Butterworths Scientific Publications).
Krebs, H., 1951, *Semi-conducting Materials*, ed. H. K. Henisch (London: Butterworths Scientific Publications).
Mellor, J. W., 1923, *Inorganic and Theoretical Chemistry* (London: Longmans, Green).
Moss, R. J., 1877, *Phil. Mag.*, **3**, 67.
Tisdale, W. E., 1918, *Phys. Rev.*, **12**, 325.
Williams, A. L., and Thompson, L. E., 1941, *J. Instn. Elect. Engrs.*, **88**, 353.
Wyckoff, R. W. G., 1931, *The Structure of Crystals* (New York: Chemical Catalog Co.).

Impurity Scattering in Oxide Semiconductors

By E. W. J. MITCHELL
Metropolitan–Vickers Electrical Co. Ltd., Trafford Park, Manchester

Communicated by B. G. Churcher ; MS received 21st August 1951

ABSTRACT. The oxide semiconductors are characterized by relatively high concentrations of impurity centres and low electron mobility. It is suggested that electron scattering by neutral impurity centres makes an important contribution to the resistivity of these materials. The familiar relation $\log \sigma_0 = \alpha + \beta\epsilon$ together with the theory given by Busch are discussed and an alternative explanation involving impurity scattering and a dependence of ϵ on N is given. This is compared with experimental results for several oxides, including measurements by the author on the $Fe_2(Ti)O_3$ system.

§ 1. INTRODUCTION

One of the features which appears to differentiate the oxide semiconductors from those, such as germanium and silicon, in which the binding is covalent is the high concentration of impurities* required in the former to provide conductivities of the order of 10^{-1} or 10^{-2} $(\Omega\ \mathrm{cm})^{-1}$. This is illustrated in the table, where concentrations and conductivities are given for three oxides and for germanium.

	Type of impurity	Lowest conductivity $(\Omega\ \mathrm{cm})^{-1}$ $N=0$	Impurity concentration	$\sigma(\Omega\ \mathrm{cm})^{-1}$
Nickel oxide	Substituted lithium	$10^{-7}-10^{-8}$	$6{\cdot}3\times10^{21}$	10^{-1}
Ferric oxide	Substituted titanium	$10^{-8}-10^{-9}$	$7{\cdot}8\times10^{20}$	10^{-2}
Magnesium titanate	Excess titanium	10^{-14}	$4{\cdot}2\times10^{21}$	5×10^{-2}
Germanium	Substituted aluminium	$10^{-1}-10^{-2}$	6×10^{19}	2×10^{2}

---

* Throughout this paper the word 'impurity' has been used to describe lattice defects producing electronic conductivity, irrespective of whether foreign atoms are involved. Although this is not entirely satisfactory, it leads to the accepted term 'impurity scattering'.

We may expect, therefore, that phenomena associated with high impurity concentrations, which become significant in germanium only at low temperatures, will, for metallic oxides, produce measurable effects at higher temperatures. For such high concentrations the following phenomena may become important:

A. *Interaction between impurities.* This will give rise to a band of impurity levels in place of the single level which is present at lower concentrations. The width of these impurity bands will be expected to increase with N, and consequently the activation energy relevant to the conduction process will decrease. It also allows of the possibility of conduction by electron transfer in the impurity band.

B. *Scattering of the conduction electrons* by the impurity atoms, either (i) by ionized centres, or (ii) by neutral centres.

The marked decrease of ϵ with increasing N which is found for those oxide semiconductors for which values of N have been determined indicates that interaction is appreciable in these substances; fig. 1 shows results for nickel oxide and ferric oxide containing substituted lithium and titanium respectively.

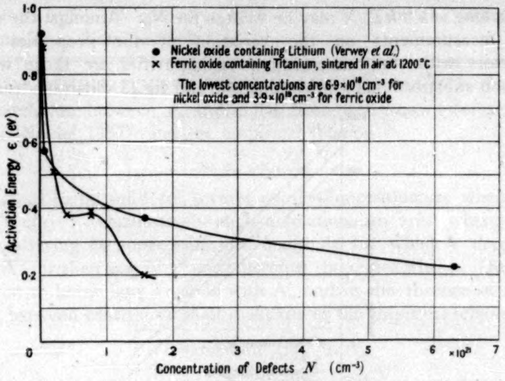


Fig. 1. Variation of activation energy with concentration of defects.

The activation energies commonly found for oxides, even allowing those with high concentrations of impurity, are such that the number of neutral centres at room temperature will always be very much greater than the number ionized, so that if impurity scattering plays an important part it must be of type B (ii), that is, scattering from neutral centres. The main purpose of this note is to examine this possibility and make quantitative comparisons with experimental work.

§ 2. ELECTRON MOBILITY AND NEUTRAL IMPURITY SCATTERING

All the phenomena mentioned above can be thought of in terms of the analogy between hydrogen atoms and ionic defects in electronic semiconductors (Mott 1940, 1951). Thus just as energy bands are formed when the distance between neighbouring atoms decreases, so impurity bands will be formed in semiconductors, the impurities being considered as hydrogen atoms embedded in a medium of dielectric constant κ. The two cases of impurity scattering then correspond to electron scattering by hydrogen nuclei and to scattering by neutral hydrogen atoms. It must be borne in mind that this analogy applies to semiconductors containing singly charged centres such as germanium with substituted aluminium, nickel oxide with lithium; when the defect is doubly charged as, for example, in reduced titanium dioxide, in oxidized nickel oxide or in titanium dioxide with substituted tantalum, the analogy is with helium atoms.

Using this model and making use of the theory of Massey and Moiseiwitsch (1950) on collisions of slow electrons and hydrogen atoms, Erginsoy (1950) finds that

$$\sigma^{-1}=4\cdot 35\times 10^{-4}\ \kappa N_{\mathrm{n}}/n\ (\Omega\,\mathrm{cm}), \qquad \ldots\ldots(1)$$

where N_{n} is the concentration of neutral centres, n is the concentration of electrons in the conduction band, κ is the dielectric constant and σ is the conductivity. Thus the mobility v_{n} is given by

$$v_{\mathrm{n}}=\frac{\sigma}{n\boldsymbol{e}}=\frac{1\cdot 47\times 10^{22}}{\kappa N_{\mathrm{n}}}\ \mathrm{cm}^2/\mathrm{volt}/\mathrm{sec}, \qquad \ldots\ldots(2)$$

in which, as long as $\epsilon \gg \boldsymbol{k}T$, N may be written for N_{n}. Amongst the most easily controlled measurements on the variation of the properties of oxide semiconductors with the concentration of impurities are those which have been made on substitutional semiconductors, and fig. 2 illustrates the variation of v with N for two such systems. These values are calculated from the conductivity measurements of Verwey *et al.* on nickel oxide containing various amounts of lithium and from our own measurements on ferric oxide containing titanium. It has been assumed that the Boltzmann statistics may be used to describe the carrier distribution so that the mobility v can be determined from a knowledge of σ, N and ϵ for each sample: $\sigma=\boldsymbol{e}vAN^{1/2}\exp(-\epsilon/2\boldsymbol{k}T)$ or

$$\sigma_0=\boldsymbol{e}vAN^{1/2} \quad \text{and} \quad v=\sigma_0/\boldsymbol{e}AN^{1/2}. \qquad \ldots\ldots(3)$$

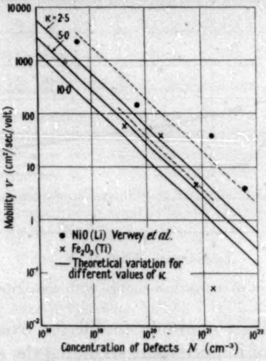


Fig. 2. Variation of mobility with concentration of defects, for nickel oxide–lithium and ferric oxide–titanium.

The results plotted in fig. 2, while not agreeing with (2) for reasonable values of κ, do indicate a regular variation of v with N which can be represented approximately by $\log v = \text{constant} - \log N$, or $v \propto 1/N$. This form of variation of mobility with concentration can only arise, it seems, from electron scattering by neutral impurity centres, for scattering by lattice ions provides a mobility which is independent of the concentration of defects, while it is unlikely that scattering by ionized impurity centres, which does involve a dependence of mobility on concentration, is significant in the present samples because of the relatively small amount of ionization. Thus in the case of the ferric oxide specimens the number of ionized centres, again assuming Boltzmann statistics to be applicable, varies from $5{\cdot}1 \times 10^{12}$ to $2{\cdot}5 \times 10^{18}$ cm^{-3}, while the total concentration of defects is increased from $3{\cdot}92 \times 10^{19}$ to $1{\cdot}57 \times 10^{21}$ cm^{-3}. The high values of mobility calculated from scattering of the electrons by these ionized impurities show that we can neglect such scattering. For example, using the Conwell–Weisskopf (1950) relation with $\kappa = 5$,

$$v_{\mathrm{I}} = \frac{4{\cdot}1 \times 10^{20}}{N_{\mathrm{I}}} \frac{1}{\ln \{1 + 6{\cdot}75 \times 10^{12}/N_{\mathrm{I}}{}^{2/3}\}}$$

where N_{I} is the concentration of ionized centres, we find values of v_{I} ranging from 8×10^6 to $1{\cdot}05 \times 10^2$, compared with values of v_{N}, for the same conditions, ranging from 70 to 2 cm^2/sec/volt.

§ 3. NEUTRAL IMPURITY SCATTERING AND THE RELATION $\log \sigma_0 = \alpha + \beta\epsilon$

The correlation between σ_0 and ϵ has been well known for several years (Meyer and Neldel 1937), and the empirical relation

$$\ln \sigma_0 = \alpha + \beta\epsilon \qquad \ldots\ldots(4)$$

has been found to be valid for several oxide semiconductors when σ_0 is less than $10^{-2}\,(\Omega\,\mathrm{cm})^{-1}$. Qualitatively such a relation can arise whenever neutral impurity scattering is appreciable in materials for which ϵ decreases with increasing N, for then $\sigma_0 \propto N^{-1/2}$ and, therefore, increases with ϵ. Quantitatively we first need to know how ϵ varies with N, and in the absence of a theory of interaction between centres we shall make use of the empirical relation indicated by fig. 3:

$$\epsilon = \text{constant} - B^{-1} \ln N, \qquad \ldots\ldots(5)$$

i.e. $N = N_0 \exp(-B\epsilon)$ or $l = l_0 \exp(-B/3\epsilon)$, where l is the average distance between centres. (This variation of ϵ with N is of a form quite different from that found for silicon; thus Bardeen and Pearson (1949) represented their results by $\epsilon = \epsilon_0 - B'/l$.) For nickel and ferric oxides we find, from fig. 3, values of B of 10·4 and 5·3 respectively.

On the basis of electron scattering from neutral impurity centres we have from (2) and (3) and for the constant $A = 3{\cdot}2 \times 10^9$ at room temperature $\sigma_0 = 7{\cdot}91 \times 10^{12} \kappa^{-1} N^{-1/2}\,(\Omega\,\mathrm{cm})^{-1}$, so that for $\kappa = 5$ we have

$$\log \sigma_0 = 12{\cdot}20 - \tfrac{1}{2} \log N. \qquad \ldots\ldots(6)$$

Combining this with (5),

$$\log \sigma_0 = (12{\cdot}20 - \tfrac{1}{2} \log N_0) + B\epsilon/4{\cdot}6. \qquad \ldots\ldots(7)$$

This is the Meyer and Neldel relation, and from their results* we can deduce values of B for oxide systems in which the dependence of activation energy on N has not been studied.

---

* The original values of ϵ used by Meyer and Neldel were based on $\sigma = \sigma_0 \exp(-\epsilon/\boldsymbol{k}T)$; these values of B refer to their results expressed in terms of $\sigma = \sigma_0 \exp(-\epsilon/2\boldsymbol{k}T)$.

The values of B deduced in this way from Meyer and Neldel's work on ferric oxide, titanium dioxide and zinc oxide are 13·8, 8·2 and 11·5 respectively. For ferric oxide we have seen that there is a discrepancy of a factor of 2·4. This lack of agreement is disappointing, for the values of B deduced from Meyer and Neldel's results are not dependent on quantitative agreement between experimental and theoretical values of σ_0, but only on the form of their variation with N. Scattering from grain boundaries, which may be important in polycrystalline materials, and which would affect the values of σ_C, can account for the discrepancy only if the distribution of boundaries is dependent on N.

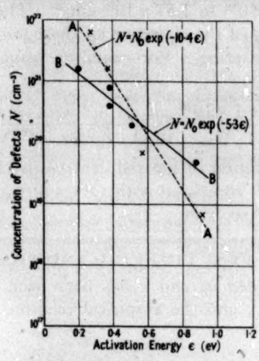


Fig. 3. Variation of activation energy with concentration of defects: A for nickel oxide–lithium, B for ferric oxide–titanium.

Alternatively, the difference may be connected with the different nature of the defects in the two types of semiconductor. In both sets of results the activation energy is varying with N: in the one case this variation has been determined experimentally, while in the other—that is for oxidation or reduction semiconductors—it has been assumed to be of the same form. The centres in the substitutional semiconductors which we have considered may be regarded as analogous to hydrogen atoms, while in the case of oxides made conducting by oxidation or reduction the defects should be considered as helium atoms. To a first approximation, and in the absence of interaction between centres, the outer electron is bound to the immobile positive charge with the same energy in both cases; but it would be reasonable to expect that when the defects are sufficiently concentrated for interaction to occur this energy would not be the same and, even if both gave a variation of the form of (5), that the values of B would be different.

§4. DISCUSSION

Busch (1946) has already given a quantitative explanation of (4), which depends on three assumptions: (i) that there is a distribution of activation energies depending upon a characteristic freezing-in temperature θ, a constant for a given material; (ii) that for specimens containing larger amounts of impurity the effective activation energy is *increased*; (iii) that lattice scattering predominates. It seems unlikely, from the results described in this paper, that assumptions (ii) and (iii) are applicable to the specimens used by Meyer and Neldel, at least to those for which $\sigma_0 < 10^{-2}\,(\Omega\,\mathrm{cm})^{-1}$ and for which the relation (4) was found. In the absence of the type of interaction discussed in §1, and under certain conditions when impurity scattering is negligible, there would seem to be two mechanisms from which the assumptions made by Busch might arise. The first requires the existence of cations and anions in different types of lattice position, that is positions characterized by different configurations of neighbours. For example, in titanium dioxide (rutile) there are two types of cation site, those at the corners of the unit tetragon and those at the body centres. This can give rise to different activation energies as illustrated in fig. 5. On the other hand, for semiconductors having a rocksalt structure, e.g. nickel oxide (fig. 4) there is only one type of cation position. In this semiconducting oxide whether the conduction results from oxidation, (4 *a*) or substitution (4 *b*), the readjustment of charge may be considered as the formation of Ni^3 ions in positions adjacent to the defect, and for this structure all such positions are identical. The process requiring an activation energy is the transfer of an electron from a non-adjacent Ni^2 to a Ni^3 ion adjacent to the defect, for once this has taken place the electron can move randomly, by electron transfer, throughout the lattice. It follows from the identity of cation sites that this process can take place in only one way, thus leading to a unique activation energy.

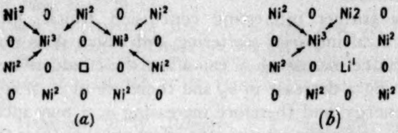


Fig. 4. (*a*) Nickel oxide with excess oxygen. (*b*) Nickel oxide with lithium.

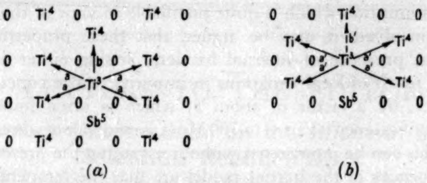


Fig. 5. Semiconducting titanium dioxide illustrating processes requiring different activation energy: (*a*) Sb^5 ion in body centre; electron released from defect by transfers such as a or b; (*b*) Sb^5 ion in corner position; electron released from defect by transfers such as a′ or b′.

Alternatively there might exist in the lattice, as well as single impurity centres, isolated pairs of impurity centres or impurities in even higher stages of aggregation. Electrons would be bound to these with different energies, and differences in the effective activation could arise if the method of preparation favoured the production of pairs rather than isolated centres. Such a process could apply to symmetrical structures like nickel oxide as well as to substances of less symmetrical structure. Certainly, however, in the present work, and probably in Meyer and Neldel's, effects of this sort are masked by the strong interaction between centres and the rapid *fall* of activation energy with increasing concentration. Moreover, N is so large that impurity scattering is important; whereupon relation (4) follows immediately.

The treatment given in this paper will not be expected to hold for activation energies less than about 0·1 ev, for under these conditions the concentration of free carriers will be such that the Boltzmann statistics can no longer be applied, while the scattering of carriers by the comparable concentration of ionized impurity centres will lead to an appreciable contribution to the mobility. Thus Meyer and Neldel (1937) found for values of room temperature conductivity greater than $10^{-2}\,(\Omega\,\mathrm{cm})^{-1}$ and activation energies less than 0·1 ev that eqn. (4), valid for lower conductivity specimens, did not apply, and that log σ_0 was almost independent of ϵ.

There is one further interesting conclusion which follows from the assumption of neutral impurity scattering and which does not depend on the statistics used. Since increasing N can affect the conductivity in two contrary ways: (*a*) by causing a decrease in σ_0 and therefore in σ, or (*b*) by producing a lower activation energy, and therefore increasing σ, σ may increase or decrease as σ_0 decreases, depending on whether the effect of (*a*) or (*b*) is the greater.

Although we have seen that several of the familiar properties of oxide semiconductors can be accounted for by the assumption of neutral impurity scattering, an assumption which is quite justifiable in view of the large impurity concentrations involved, it may be argued that these properties can also be attributed to the presence of internal barriers, arising either at intergranular contacts or as a result of local variations in impurity centre concentration. The fall of resistance, by a factor of about 5, which is often found for sintered specimens as the frequency is raised to 100 Mc/s, is evidence of some heterogeneous structure, but this can be interpreted without assuming the presence of barriers. Further consequences of the barrier model are that the temperature coefficient of resistance will fall as the frequency is increased and that the d.c. value of σ_0 will increase with increasing concentration. For the nickel and ferric oxide specimens whose properties we have been discussing the variation of σ_0 was in the opposite direction, while Volger (1951) has found, for strontium manganites containing substituted lanthanum, that the temperature coefficient is independent of frequency. In addition, the single-crystal measurements which have been made (Busch 1950, Breckenridge and Hosler 1951) have shown the same type of variation of activation energy with concentration and, in the latter case, low mobility, as is found for sintered materials.

§ 5. CONCLUSION

The calculations of mobility values from the theory of neutral impurity scattering give results which, though not in exact numerical agreement with the experimental results, do suggest that this process must be taken into account when analysing results obtained with oxide semiconductors. It is true that the experimental results with which the present comparisons are made referred to sintered specimens and it is quite conceivable that the values of σ_0 used are not entirely characteristic of the homogeneous structure which has been assumed. However, Breckenridge and Hosler (1951) have recently carried out measurements on both single crystals and dense sintered specimens of titanium dioxide and found little difference in their electrical properties; values of mobility, for example, were in both cases low.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. W. Sillars for many discussions on these problems, Dr. C. Dannatt, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan–Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

BARDEEN, J., and PEARSON, G. L., 1949, *Phys. Rev.*, **75**, 865.

BRECKENRIDGE, R., G., and HOSLER, W. R., 1951, *Amer. Phys. Soc. Bull.*, April.

BUSCH, G., 1946, *Helv. phys. Acta*, **19**, 189 ; 1950, *Ibid.*, **23**, 530.

CONWELL, E., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77**, 388.

ERGINSOY, C., 1950, *Phys. Rev.*, **79**, 1013.

MASSEY, H. S. W., and MOISEIWITSCH, B. L., 1950, *Phys. Rev.*, **78**, 180.

MEYER, W., and NELDEL, H., 1937, *J. Tech. Phys.*, **18**, 588.

MOTT, N. F., 1940, *Electronic Processes in Ionic Crystals* (Oxford: University Press); 1951, *Semi-Conducting Materials*, ed. H. K. Henisch (London: Butterworths Scientfiic Publications).

VERWEY, E. J. W., HAAIJMAN, P. W., ROMEIJN, F. C., and OOSTERHOUT, G. W., 1950, *Philips' Res. Rep.*, **5**, 173.

VOLGER, J., 1951, *Semi-Conducting Materials*, ed. H. K. Henisch (London: Butterworths Scientific Publications.)

LETTERS TO THE EDITOR

Some Further Observations on the Effect of Bending on Selenium Rectifier Discs

It has been observed (Selényi 1947) that bending of a selenium rectifier disc causes an increase in the back current. The greater part of this back current is only temporary and disappears after the bending has been continued for a few minutes. However, a small part persists and the behaviour of the disc in this state is that of a polarized system. Also, an increase of the bending causes the back current to increase and vice versa. Since the effect is capable of indefinite repetition, within certain limits, it could be used for the conversion of acoustical or mechanical vibrations into corresponding alternating currents. This possibility was explored in the following improvised experiment. From the middle part of a rectifier disc of 55 mm diameter a strip, about 1 cm in width, was cut. With both ends secured the strip was then permanently bent by a spiral spring attached to its centre. A reverse voltage of 10 v was applied to the rectifier strip and connection made through a three-stage low-frequency amplifier to a loudspeaker. On striking the spiral spring its tone was clearly heard in the loudspeaker, from which it was apparent that the frequency range of the bending effect extends to at least some hundred cycles per second.

This observation, as well as those enumerated in my first communication, were made on normally manufactured discs, that is on discs which had been finished by the 'electric forming' process. This process consists in loading the disc with successively increased voltages and for a longer time in the reverse direction, which causes successive ionization of the impurity centres located in the blocking layer and wandering of the ions thus produced (Selényi 1943). The bending effect has also been ascribed by me to the ionization of the impurity centres under the simultaneous action of the strong electric field in the blocking layer and the elastic stresses impressed on the selenium layer by the bending. Consequently, the electric forming—that is the partial removal of the impurity centres from the blocking layer—must have a strong influence on the bending effect. This was substantiated by the bending with the hand of a sample of electrically unformed disc, which produced at the low figure of 3 v a 50% increase in the back current. Forming for some minutes diminished the effect to 10%, and after perfect forming the effect of bending at the same low voltage became just observable. Similar results have been obtained on many other discs. The electric forming being the vital process in manufacture, the detailed study of the bending effect is not only of theoretical interest but will, I think, prove useful in the existing testing and controlling of selenium rectifier manufacture.

Physical Institute,
Roland Eötvös University,
Budapest VIII.
8th October 1951.

P. Selényi.

Selényi, P., 1943, *Elecktrotech. u. Maschinenb.* [*E.u.M.*], **61,** 633; 1947, *Nature, Lond.*, **160,** 197.

The Approach to Saturation Magnetostriction of Nickel

By assuming that near saturation the magnetization of a ferromagnetic material proceeds wholly by domain vector rotations against the magneto-crystalline forces, Rüdiger and Schlechtweg (1941) have shown that in fields sufficiently great to approach saturation the magnetostriction should obey a law of the form

$$dl/l=(dl/l)_s-a/H, \qquad \ldots\ldots(1)$$

where $(dl/l)_s$ is the saturation magnetostriction and a is a constant. Their analysis shows the form of this equation is independent of the direction of easy magnetization in the material. The experimental results quoted by them for iron, both in single crystal (Honda and Masiyama 1926) and polycrystalline forms (Kornetzki 1933) are not in very good agreement with eqn. (1). This may be due to the experimental difficulty of making accurate measurements in a region where dl/l is varying slowly with H or it may indicate genuine departures from eqn. (1).

Now it is known that in this region the magnetization follows a law of approach expressed by

$$I=I_s(1-c/H-b/H^2), \qquad \ldots\ldots(2)$$

in which c and b are constants such that $c\ll b$, so that the term in $1/H$ is operative only in fields above a few thousand oersteds. Now since the magnetostriction near saturation is approximately proportional to I^2, eqns. (1) and (2) appear to be incompatible.

In order to test Rüdiger and Schlechtweg's conclusions, measurements have been made of the differential magnetostriction $(dl/dH)/l$ as a function of a steady applied field. The material chosen was pure nickel rather than iron, because (*a*) the magnetostriction coefficient is somewhat greater than in iron, and (*b*) magneto-crystalline energy and the fields required to produce saturation are smaller. The specimen used was in the form of a strip and was annealed *in vacuo* by maintaining it at 820° c for four hours and cooling at a rate of 100° c an hour. Measurements were made using resistance strain gauges and a photoelectric galvanometer amplifier.

The experimental results given in the accompanying figures show that eqn. (1) is in general not obeyed, but that $(dl/dH)/l$ is a linear function of $1/H^3$ over a considerable range of H. The range of H over which $(dl/dH)/l$ is proportional to $1/H^2$ (fig. 1) is thought to be too small to be significant.

The straight line of fig. 2 may be represented by the equation

$$\frac{1}{l}\frac{dl}{dH}=\frac{\alpha}{H^3}+\beta.$$

This may be integrated, giving

$$\frac{dl}{l}=-\frac{\alpha}{2H^2}+\beta H+\gamma, \qquad \ldots\ldots(3)$$

where γ is an integration constant. Since β is very small ($\sim10^{-9}$), we may, to a good approximation, neglect the second term, and then γ may be put equal to the saturation magnetostriction. Equation (3) may therefore be written

$$\frac{dl}{l}=\left(\frac{dl}{l}\right)_s\left(1-\frac{\alpha'}{H^2}\right)+\beta H, \qquad \ldots\ldots(4)$$

in which α' is a constant and β a measure of the field dependence of the spontaneous magnetostriction, the latter having a value given by the intercept on the graph in fig. 2 of about $1{\cdot}3 \times 10^{-9}$ per oersted. From an investigation of the volume magnetostriction in high fields Döring (1936) found that the change in spontaneous magnetostriction with field is small, the quantity $(dl/dH)/l$ being about 3×10^{-11} per oersted for nickel at room temperature. This fact, together with our observed value of β, may be taken to indicate that the extrapolation of the straight line of fig. 2 to $H = \infty$ cannot be made, and in very high fields a linear variation of dl/l with $1/H$ may occur as predicted.

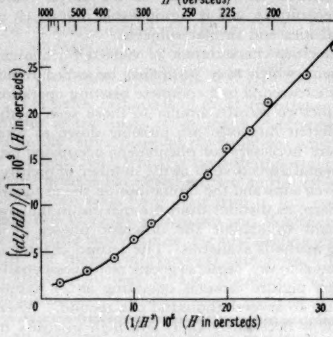


Fig. 1.

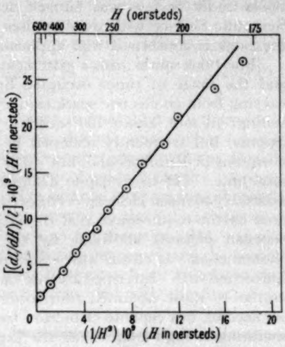


Fig. 2.

A fuller account of this work, together with some theoretical justification for eqn. (4), will be given in a later paper.

The University,
Nottingham.
6th December 1951.

E. W. Lee.

Döring, W., 1936., *Z. Phys.*, **103**, 560.
Honda, K., and Masiyama, Y., 1926, *Sci. Rep. Tohoku Imp. Univ.*, **15**, 755.
Kornetzki, M., 1933, *Z. Phys.*, **87**, 576.
Rüdiger O., and Schlechtweg, H., 1941, *Ann. Phys., Lpz.*, **39**, 1.

REVIEWS OF BOOKS

A Text Book of Physics: Part IV—Sound, by J. Duncan and S. G. Starling. Pp. xi + 105. 2nd Edition. (London: Macmillan, 1950.) 5*s*.

The Sound section of this well-known textbook now appears in a separate issue, and as such should be invaluable to those students whose work for the Intermediate Science or equivalent examination has not included this subject, still much neglected in schools. The presentation of the material is interesting, the diagrams clear and the well-chosen exercises help the student to test his knowledge. Among new features are paragraphs on the decibel scale and ultrasonics.

E. G. R.

The Photographic Study of Rapid Events, by W. D. CHESTERMAN. Pp. xiii + 167. (Oxford: Clarendon Press, 1951.) 21*s*.

The last war saw great use made of photographic methods in the study of events which occur too rapidly for their precise nature to be appreciated by the eye. The author of the book under review was himself actively concerned as a member of the Royal Naval Scientific Service with war-like uses of these techniques, and it is noticeable that much of the book is concerned with applications to ballistics and similar subjects.

The book starts with a statement of the velocities characteristic of various phenomena and the range of times occupied by phenomena which may justifiably be called rapid, varying from an electric spark lasting about a microsecond to a complete blasting operation lasting, all told, about 100 seconds. Motion picture records similar to those seen in the cinema, but frequently obtained by quite different methods, are possible down to total durations of 1 millisecond, but other methods are necessary for phenomena occupying less total time. The first chapter discusses fundamental aspects such as the number of pictures necessary to make clear an event occupying a given time and the limitations on the exposure time by the requirement that the picture be sharp, as distinct from the maximum interval between pictures allowable by the requirement to exhibit the different phases of a phenomenon. It closes with a summary of the methods available. The second chapter is concerned with ' Intermediate rate cameras ', that is to say, cameras giving results essentially similar to those obtained from ordinary motion picture cameras operating at 24 frames per second, but capable of picture frequencies up to several thousand per second. When operating at high frequencies the exposure time is necessarily short and high intensity of illumination is necessary. The third chapter deals with methods of illumination. This is followed in Chapter IV by the closely related subject of ' Choice of sensitive material '. The following chapters then deal with ' Single pictures ', ' Film drum cameras ', ' Spark and schlieren photography '. This concludes the first part of the book on the technique. Part II deals with applications, under the headings Zoological studies, Biological and medical sciences, Physical and engineering research and Military applications. In addition to 42 diagrams in the text, there are 32 half-tone plates, with a total of nearly 300 individual pictures. Some of these latter require careful study to discover the details of the phenomena they record.

Compression of all the above into little more than 150 pages of text means that only a broad outline is attempted. This is reasonable enough, in the first book published in English on its subject as far as the reviewer is aware, but it does mean that it is more suited to the ' general reader ' or the beginner than to one engaged in high speed photography. The broad treatment also tends to rather loose writing in places and to omissions. In the first chapter there are no less than five diagrams relating exposure time, event velocity, optical magnification and permissible image movement, which seems unnecessary elaboration on a simple theme. The important problem of getting as much light as possible on the subject, from a given lamp, is dismissed in the sentence " Techniques are now established for the efficient use of extended sources of illumination of reasonably large areas, and also for the concentration of high light intensities on a limited surface under study ". Some excursion into the principles of illumination engineering and description of methods would be very helpful. It is not easy sometimes, for example, to decide whether to use a bare lamp close to the subject or to use a projector.

The chapter on Choice of sensitive materials makes no mention of the types of material available, nor of development and processing generally. The text, in fact, bears little relation to the title, more space being devoted to the characteristics of flashtubes than to photographic materials. The sort of question one wants an answer to is how materials are affected by processing (including intensification, perhaps), whether to use the fastest materials one can get (usually at the expense of definition in the photographs), or whether one could manage with a slower material, of finer grain. These questions are admittedly difficult to answer satisfactorily in general terms, but some helpful discussion of them ought to be possible, if only by giving examples of good photographic technique.

The treatment of spark and schlieren photography is very elementary, and one would like to see more extended discussion to include the effects of size of light source and diffraction, not necessarily by elaborate mathematical analysis, but possibly by a series of photographs taken under different conditions. These questions are not unimportant, for on the answer to them depends the optical quality of the mirrors or lenses to be used in schlieren methods.

The above comments are offered in the hope that they may be useful to the author, for, as it stands, the book is pleasantly written and produced in the excellent fashion one has come to expect from the Clarendon Press, the information in it is sound, and it can be recommended, particularly to the non-expert. E. W. H. SELWYN.

The Oxide-Coated Cathode, Vol. I: Manufacture, by G. HERRMANN and S. WAGENER (translated by S. WAGENER). Pp. viii + 148. (London: Chapman and Hall, 1951.) 21*s*.

The oxide cathode is of great importance technically, since its operation as a thermionic emitter is the basis of so many electronic devices. It is also of considerable interest physically, since many of the problems which it introduces are concerned with the physics of the solid state, which is such an important and active branch of modern physics. Several review articles have been written on the subject of the oxide cathode, and sections of certain well-known texts have dealt with it, but no book wholly devoted to the subject appeared until this work by Herrmann and Wagener was published in Germany in 1944. It was divided into two volumes, the first dealing with the manufacturing processes, and the second with the underlying physical principles. These volumes have now been translated and brought up to date by Dr. Wagener, and this review is concerned with the first of the volumes.

.B
5
2

After a brief historical survey and a description of the different types of cathode, there is an excellent account of the problems connected with the choice of core metal, the preparation of the carbonate suspensions and their application to the core metal. The production of insulated heaters for indirectly heated cathodes is included in this section. A later section describes the treatments given while pumping and the subsequent processes for developing the electron emission, and continues with an account of emission measurement, typical values of emission obtainable and other cathode characteristics. These include in particular the life, and various factors which can lead to poor performance. The final sections describe various special forms of oxide cathode.

The book is well produced and illustrated and the information is authoritative and up to date. The bibliography is excellent. The translation is good and the style very readable. In the few places where there are deviations from normal English practice, there is no loss in clarity. The only exception noticed is one rather obscure sentence on p. 13, beginning with " The vapour pressure . . . .". The proof-reading has been good, and the book can be thoroughly recommended to all who are interested in the production and use of oxide cathodes. D. A. W.

The Oxide-Coated Cathode, Vol. II: Physics, by G. HERRMANN and S. WAGENER (translated by S. WAGENER). Pp. xiv + 302. (London: Chapman and Hall, 1951.) 42*s*.

Among those who are concerned with semiconductors and their applications, the oxide-coated cathode is often regarded as a branch of solid state physics, so highly specialized and so firmly in the hands of experts as to discourage the attentions of the interested outsider. The publication of an extensive and lucid review which brings the subject back into the sphere of more general discussions is therefore particularly welcome and will be generally appreciated. The present volume is based on a work first published in Germany in 1944. It has since been brought up to date by the inclusion of more recent investigations, and the literature has been covered up to the beginning of 1950.

In order to discuss the complicated phenomena associated with oxide-coated cathodes it is necessary to give a fairly detailed description of the corresponding phenomena observed on metals and of the experimental techniques developed for the investigation of emission processes. Similarly, it is necessary to give some theoretical account of electronic as well as ionic conduction in non-metallic solids, with special reference to oxides. A lengthy introduction is thus inevitable and, indeed, about half the volume of the book is devoted to these matters. The second half deals with such topics as the properties of the oxide coating as an excess semiconductor, the interpretation of emission phenomena in terms of the electronic energy relations within the solid and at its surface, experimental methods for the determination of the work function, the mechanism of the activation process, the properties of mixed oxide coatings, the conditions governing electrical stability, and many other important problems. There is also a short Appendix on current fluctuations. An extensive system of references to original publications will make this account valuable for those who are already working in this field. The simplifying assumptions which it is from time to time convenient to make are clearly stated throughout these specialized discussions. They are later relaxed in conformity with the more complicated conditions encountered in practice. The consistency with which this is done is a valuable feature of the book.

From the theoretical point of view there are some topics of which an even more extensive treatment would have been welcome, e.g. the effect of a non-uniform distribution of activating centres on the energy relations within the oxide coating, impurity scattering and its effect on the conduction properties of the cathode, the energy relations at the contact between core metal and oxide coating, the experimental evidence (if any) for the existence of a band of impurity levels and of electron traps as suggested by Nijboer, and the role of surface levels in the mechanism of cathode poisoning. The short account of Meyer's rule is in need of some amendments, to allow for the characteristically different behaviour of solids with high and low impurity content respectively.

The book as a whole leaves the reader conscious of the substantial advances which have been made during recent years (many by the present authors), and also of the fact that the oxide-coated cathode will continue to be a subject of investigations for a long time to come.

H. K. H.

CONTENTS FOR SECTION A

PAGE
Mr. Behram Kurşunoğlu. Einstein's Unified Field Theory 81
Dr. G. J. Kynch. The Two-Body Scattering Problem with Non-Central Forces: I—Non-Relativistic 83
Dr. G. J. Kynch. The Two-Body Scattering Problem with Non-Central Forces: II—Relativistic 94
Dr. F. Mandl and Dr. T. H. R. Skyrme. Analysis of High-Energy Neutron Production 101
Dr. M. Smollett. The Frequency Spectrum of a Two-Dimensional Ionic Lattice 109
Dr. T. E. Nevin, Dr. M. Conway and Mr. M. Cranley. The Rotational Analysis of the (0,0) and (0,1) Bands of the $^7\Pi \rightarrow {}^7\Sigma$ System of Manganese Deuteride . 115
Dr. R. A. Durie. The Excitation and Intensity Distribution of CH Bands in Flames 125
Prof. L. F. Bates and Dr. C. D. Mee. The Domain Structure of a Silicon–Iron Crystal 129
Prof. L. F. Bates and Dr. C. D. Mee. A Study of Surface Closure Domains by the Powder Pattern Technique 140

Letters to the Editor:

Dr. L. R. B. Elton and Dr. H. H. Robertson. On the Radiative Correction to the Coulomb Scattering of Electrons 145
Dr. T. S. Moss. Properties of Films of Non-Metallic Antimony . . 147
Mr. G. Drummond and Dr. R. F. Barrow. The Dissociation Energy of SnO 148
Dr. K. W. H. Stevens. Line Widths in Ferromagnetic Resonance . . 149
Dr. P. J. Grant. Some Regularities in Nuclear Energy Levels . . . 150
Mr. A. Gatterer and Mr. S. G. Krishnamurty. The Spectrum of Lutecium Monoxide 151
Miss N. D. Sayers. Meinel's Infra-Red Auroral Bands 152
Reviews of Books 154
Contents for Section B 158
Abstracts for Section B 159

ABSTRACTS FOR SECTION A

Einstein's Unified Field Theory, by Behram Kurşunoğlu.

ABSTRACT. In this paper it is shown that, under certain assumptions about the metric of the space–time, energy momentum tensor of the total field, when Einstein's field equations are satisfied, vanishes identically.

A new version of the unified field theory is suggested, and it is indicated briefly that a genuine energy momentum tensor exists which is conserved and has a structure similar to that of general relativity.

From the form of the energy momentum tensor and from the field equations of the proposed theory it is concluded that the law of force for a charged mass particle moving in a gravitational and electromagnetic field would have the expected form.

The Two-Body Scattering Problem with Non-Central Forces : I—Non-Relativistic, by G. J. Kynch.

ABSTRACT. A perturbation theory is developed for the two-body problem so that central and non-central forces can be treated in a consistent manner. A phase matrix $S(r)$ is introduced which satisfies a first-order Riccati equation. The asymptotic value of S occurs in the formula for the cross section, and for a special choice of functions is the Heisenberg matrix of the problem. The formalism leads to a simple comparison of neutron–proton and proton–proton scattering, an analysis of the degree of separation of the simultaneous equations which occur, and of the most suitable choice of $S(r)$ for calculation.

The Two-Body Scattering Problem with Non-Central Forces: II—Relativistic, by G. J. Kynch.

ABSTRACT. It has been shown that both for one particle scattered by a centre of force and in the two-body problem, treated relativistically, equations for the S matrix of the Riccati type can be found exactly similar to and as simple in form as those of the non-relativistic approximation. The analytic continuation of this matrix, to determine bound states, has the same limitations as in the non-relativistic problem.

Analysis of High-Energy Neutron Production, by F. Mandl and T. H. R. Skyrme.

ABSTRACT. The energy spectra of neutrons scattered in the forward direction when different target nuclei are bombarded with high energy protons are calculated and compared with recent measurements. The calculation is subdivided into three stages : (i) as a crude analogue to the nuclear scattering, the scattering of a single neutron in a square well is calculated using Born approximation; (ii) single scattering processes are estimated classically using the results of (i) to allow for the momentum distribution of the target neutron, and allowing for the attenuation of incident and scattered beams; (iii) multiple scattering is estimated similarly. Several simplifying assumptions have to be made and double collisions only are estimated.

The calculated differential cross sections are in fairly good agreement with experiment, but the calculated energy spectrum falls off too quickly. The effect of neglecting certain processes, such as triple collisions, is discussed, as well as the applicability of the nuclear model used.

The Frequency Spectrum of a Two-Dimensional Ionic Lattice, by M. Smollett.

ABSTRACT. The frequency spectrum of a two-dimensional ionic lattice has been evaluated. A method is given for calculating the exact density of the spectrum at special points in phase space. Two logarithmic infinities, first found by Montroll using a quasi-elastic forces model, are confirmed for an ionic lattice. The bearing of these infinities on the interpretation of the second-order Raman spectra is discussed.

The Rotational Analysis of the (0,0) *and* (0,1) *Bands of the* $^7\Pi \rightarrow {}^7\Sigma$ *System of Manganese Deuteride*, by T. E. NEVIN, M. CONWAY and M. CRANLEY.

ABSTRACT. The band spectrum of manganese deuteride has been photographed at a dispersion of 1·25 Å/mm. The heads of the (0,0) and (0,1) bands of the $^7\Pi \rightarrow {}^7\Sigma$ system are at λ5703·7 Å and λ6073·3 Å respectively. It has been possible to make a rotational analysis of the two bands. The constants of the final state have been derived from the analysis and the fine structure resulting from the interaction of spin and rotation has been studied. The $^7\Pi$ state is perturbed and the terms have been evaluated from the term expressions for the final state and the wave numbers of the P, Q and R branches. The results throughout are in good agreement with the results for the corresponding bands in the spectrum of manganese hydride.

The Excitation and Intensity Distribution of CH Bands in Flames, by R. A. DURIE.

ABSTRACT. An apparent selective excitation to the $v'=1$ level of the CH 3900 Å ($^2\Sigma^- \rightarrow {}^2\Pi$) system occurs in hydrogen flames containing small amounts of organic compounds since the (1, 0) and (1, 1) bands are abnormally strong. This is accompanied by a suppression of the usual predissociation in this system. An additional band of this system, the (1, 2) at 4494·5 Å, has been observed for the first time.

The suppression of the CH predissociation is interpreted as showing that conditions in hydrogen flames approach thermal equilibrium. The appearance of the CH predissociation in normal flames indicates that the excitation is not thermal.

The Domain Structure of a Silicon–Iron Crystal, by L. F. BATES and C. D. MEE.

ABSTRACT. Powder patterns have been obtained on the (100) and (011) surfaces of a carefully shaped single crystal of 3% silicon–iron, of dimensions 13 mm × 6 mm × 6 mm, and these have permitted a more thorough quantitative test of proposed ferromagnetic domain structures than has been made hitherto. The patterns on the (011) surface are very beautiful and complicated. Although the measurements of domain spacings on both planes under Mode III magnetization conditions, as well as measurements of the angles between the domain vectors and main domain boundaries, confirm the essential features of the main structure proposed by Néel, there are found considerable numerical discrepancies which are attributed to the complicated nature of the closure domains on the (011) surface.

A Study of Surface Closure Domains by the Powder Pattern Technique, by L. F. BATES and C. D. MEE.

ABSTRACT. The powder patterns in the neighbourhood of cavities and inclusions on the surfaces of single crystals of silicon–iron have been studied, and the photographs obtained are evidence for the essential correctness of current views on closure domain structures.

Fig. 2. Photomicrograph of a section through a selenium specimen after immersion in liquid mercury at 100° c for 100 hours.

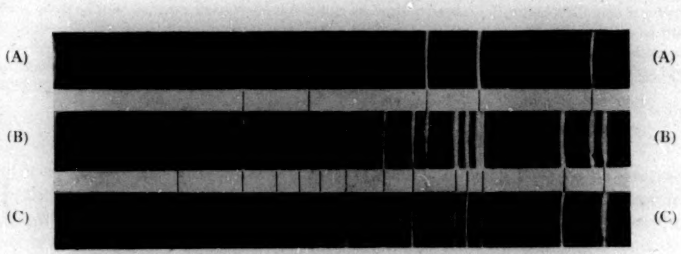


Fig. 3. X-ray powder photographs of:

(A) A surface layer detached from the specimen shown in fig. 2.
(B) Powdered crystalline selenium after exposure to mercury vapour at room temperature.
(C) Pure crystalline selenium.

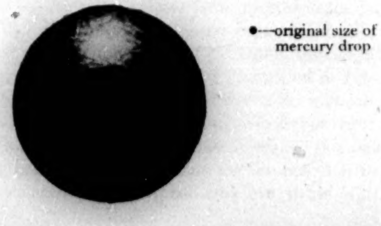


Fig. 7. Autoradiograph showing surface diffusion of mercury on selenium. A selenium rectifier disc (concentric assembly type) without counter-electrode was used for this experiment.